

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 04:55:10 ; Search time 67.53 Seconds  
(without alignments)  
4132.158 Million cell updates/sec

Title: US-09-537-654-1  
1474  
Sequence: 1 tcgaccacgcgcgcgcact.....aaaaaaaaaaaaaaaaaaaaa 1474

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCRTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfilea1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53.8	3.6	3777	3 US-09-121-321-15	Sequence 15, Appl
2	53.8	3.6	3777	4 US-08-933-803A-15	Sequence 15, Appl
3	49.8	3.4	1641	1 US-08-300-903A-8	Sequence 8, Appl
4	48.4	3.3	1824	3 US-08-606-505B-1	Sequence 1, Appl
5	48.4	3.3	1824	4 US-09-616-990-1	Sequence 1, Appl
6	47.8	3.2	958	2 US-08-757-046A-5	Sequence 5, Appl
7	47.8	3.2	958	3 US-09-447-208-5	Sequence 5, Appl
8	47.8	3.2	958	3 US-09-135-988-5	Sequence 5, Appl
9	47.8	3.2	958	4 US-09-277-716-5	Sequence 5, Appl
10	47.8	3.2	958	4 US-08-597-274A-5	Sequence 5, Appl
11	47.8	3.2	536	1 US-08-341-568-1	Sequence 1, Appl
12	47.8	3.2	536	2 US-08-811-020-1	Sequence 1, Appl
13	47.8	3.2	2852	3 US-09-027-137-2	Sequence 2, Appl
14	46.6	3.2	1386	2 US-08-897-340-1	Sequence 1, Appl
15	46.6	3.2	1386	3 US-09-252-329-1	Sequence 1, Appl
16	46.4	3.1	2447	2 US-09-014-969-14	Sequence 14, Appl
17	46.2	3.1	2296	4 US-08-496-841C-137	Sequence 137, App
18	46.2	3.1	4765	3 US-08-936-135-21	Sequence 21, Appl
19	46.2	3.1	4780	3 US-08-936-135-23	Sequence 23, Appl
20	46.2	3.1	4780	3 US-08-936-135-23	Sequence 23, Appl
21	46.2	3.1	375	3 US-08-946-026-23	Sequence 23, Appl
22	46.2	3.1	375	3 US-08-946-026-23	Sequence 23, Appl
23	45.8	3.1	1813	5 PCN-US94-12883-3	Sequence 3, Appl
24	45.8	3.1	2345	3 US-08-955-937A-1	Sequence 1, Appl
25	45.8	3.1	2345	3 US-09-300-985-1	Sequence 1, Appl
26	45.4	3.1	882	4 US-08-909-965C-9	Sequence 9, Appl
27	45.2	3.1	1858	2 US-08-909-965C-11	Sequence 11, Appl

28	45	3.1	1781	1 US-08-314-615-2	Sequence 2, Appl
29	45	3.1	1781	1 US-08-314-615-2	Sequence 2, Appl
30	45	3.1	1781	1 US-08-433-010-2	Sequence 2, Appl
31	45	3.1	1781	1 US-08-433-010-2	Sequence 2, Appl
32	45	3.1	1781	1 US-08-433-010-2	Sequence 2, Appl
33	45	3.1	1781	1 US-08-481-130-4	Sequence 4, Appl
34	45	3.1	1781	1 US-08-656-984A-4	Sequence 4, Appl
35	45	3.1	1781	1 US-08-656-984A-4	Sequence 4, Appl
36	45	3.1	1781	1 US-08-482-882-2	Sequence 2, Appl
37	45	3.1	1781	1 US-08-483-604-4	Sequence 4, Appl
38	45	3.1	1781	2 US-08-487-113D-2	Sequence 2, Appl
39	45	3.1	1781	2 US-08-487-595-4	Sequence 2, Appl
40	45	3.1	1781	2 US-08-473-503-2	Sequence 2, Appl
41	45	3.1	1781	2 US-08-483-932-2	Sequence 2, Appl
42	45	3.1	1781	2 US-08-720-420A-2	Sequence 2, Appl
43	45	3.1	1781	3 US-08-714-017-2	Sequence 2, Appl
44	45	3.1	1781	3 US-08-863-790-2	Sequence 2, Appl
45	45	3.1	1781	3 US-08-475-680-2	Sequence 2, Appl
46	45	3.1	1781	3 US-08-296-749-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-09-121-321-15  
; Sequence 15, Application US/09121321  
; Patent No. 6090783  
; GENERAL INFORMATION:  
; APPLICANT: Saiga, Akihiko  
; APPLICANT: Orita, Satoshi  
; APPLICANT: Igarashi, Hisanaga  
; APPLICANT: Okumura, Kouichi  
; APPLICANT: Sakaguchi, Gaku  
; TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF  
; TITLE OF INVENTION: GENE EXPRESSION AND NOVEL PROTEIN  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/121,321  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/933,803  
; FILING DATE: 19-SEP-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley, James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: SHGN-12C1P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3777 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 139..2151  
; US-09-121-321-15



RESULT 5  
US-09-616-990-1  
; Sequence 1, Application US/09616990  
; Patent No. 6232109  
; GENERAL INFORMATION:  
; APPLICANT: KIRUCHI, Yasuhiro  
; KIKOKAWA, Shigeto  
;

RESULT 6  
 US-08-757-046A-5  
 ; Sequence 5, Application US/08757046A  
 ; Patent No. 5876995  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bryan, Bruce  
 ; TITLE OF INVENTION: BIOLUMINESCENT ARTICLES OF  
 ; NUMBER OF SQUENCES: 14 MANUFACTURE

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Brown, Martin, Haller & McClain  
;; STREET: 1660 Union Street  
;; CITY: San Diego  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 92101-2926  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FASTSEQ Version 1.5  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/757,046A  
;; FILING DATE: 11-25-96  
;; CLASSIFICATION: 435  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/597,274  
;; FILING DATE: 02-06-96  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Seidman, Stephanie L.  
;; REGISTRATION NUMBER: 33,779  
;; REFERENCE/DOCKET NUMBER: 6680-105B  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 619-238-0999  
;; TELEFAX: 619-238-0062  
;;  
;;  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 958 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: CDNA  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; FRAGMENT TYPE:  
;; ORIGINAL SOURCE:  
;; FEATURE:  
;; NAME/KEY: Coding Sequence  
;; LOCATION: 115...702  
;; OTHER INFORMATION: apoaecorin-encoding gene  
;; PUBLICATION INFORMATION:  
;; AUTHORS: Inouye et al.  
;; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.  
;; VOLUME: 82  
;; PAGES: 3154-3158  
;; DATE: (1985)  
;; DOCUMENT NUMBER: PATENT NO.: 5,093,240  
;; US-08-757-046A-5  
;;  
Query Match 3.2%; Score 47.8; DB 2; Length 958;  
Best Local Similarity 67.7%; Pred. No. 0.00047;  
Matches 67; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
;;  
QY 1376 tagtgccttgcatacctgacatctggcccatctacatcgaagcctataatagcg9g 1435  
DB 856 TTGTTTAAATCAACAGACTTACAAATCGAAAAAGTAAAAAAGTAAAAAAGTAAAAA 915  
QY 1436 caaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 1474  
DB 916 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 954  
;;  
RESULT 7  
; Sequence 5, Application US/09447208  
; Patent No. 6113886  
; GENERAL INFORMATION:  
; APPLICANT: Bryan, Bruce  
; TITLE OF INVENTION: BIOLUMINESCENT ARTICLES OF MANUFACTURE  
; NUMBER OF SEQUENCES: 14

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Heller Ehrman White & McCauliffe  
;; STREET: 4250 Executive Square, 7th Floor  
;; CITY: La Jolla  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 92037  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FASTSEQ Version 1.5  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/447,208  
;; FILING DATE:  
;; CLASSIFICATION:  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 0909/135,988  
;; FILING DATE: 08-17-98  
;;  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/757,046  
;; FILING DATE: 11-25-96  
;;  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/597,274  
;; FILING DATE: 02-06-96  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Seidman, Stephanie L.  
;; REGISTRATION NUMBER: 33,779  
;; REFERENCE/DOCKET NUMBER: 24727-105C  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 619-450-8400  
;; TELEFAX: 619-450-8499  
;;  
;;  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 958 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: CDNA  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; FRAGMENT TYPE:  
;; ORIGINAL SOURCE:  
;; FEATURE:  
;; NAME/KEY: Coding Sequence  
;; LOCATION: 115...702  
;; OTHER INFORMATION: apoaecorin-encoding gene  
;; PUBLICATION INFORMATION: PATENT NO.: 5,093,240  
;; AUTHORS: Inouye et al.  
;; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.  
;; VOLUME: 82  
;; PAGES: 3154-3158  
;; DATE: (1985)  
;; US-09-447-208-5  
;;  
Query Match 3.2%; Score 47.8; DB 3; Length 958;  
Best Local Similarity 67.7%; Pred. No. 0.00047;  
Matches 67; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
;;  
QY 1376 tagtgccttgcatacctgacatctggcccatctacatcgaagcctataatagcg9g 1435  
DB 856 TTGTTTAAATCAACAGACTTACAAATCGAAAAAGTAAAAAAGTAAAAAAGTAAAAA 915  
QY 1436 caaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 1474  
DB 916 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 954

```

Db          916 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 954
|||||
RESULT      9
US-09-277-716-5
; Sequence 5, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryant, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLOME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; EARLIER FILING DATE: 1998-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 5
LENGTH: 958
TYPE: DNA
ORGANISM: Aequeorea (luminescent jellyfish)
FEATURE:
NAME/KEY: CDS
LOCATION: (115)..(702)
FEATURE:
OTHER INFORMATION: Apoaequorin-encoding gene
PUBLICATION INFORMATION:
PATENT DOCUMENT NUMBER: 5,093,240
PATENT FILING DATE: 1987-10-08
PUBLICATION DATE: 1992-03-03
PUBLICATION INFORMATION:
AUTHORS: Imouye, S.
TITLE: Cloning and sequence analysis of cDNA for the luminescent protein
JOURNAL: Proc. Natl. Acad. Sci. USA
VOLUME: 82(10)
PAGES: 3154-3158
DATE: 1985-05
US-09-277-716-5

Query Match                      3.2%; Score 47.8; DB 4; Length 958;
Best Local Similarity    67.7%; Pred. No. 0.00047;
Matches     67; Conservative   0; Mismatches    32; Indels    0; Gaps    0;

OY 1376 tagtgattgtatctgaacatcgtgcccatcgtaactgcagtaaagccataatagcggg 1435
||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 856 ttglttttlaatcaacagaacttacaaactcgaaaagttaaaaaaaaaaaaaaaaaaaa 915
|||||

OY 1436 caaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1474
|||||
DB 916 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 954
|||||

RESULT      10
US-08-597-274A-5
; Sequence 5, Application US/08597274A
; Patent No. 6247995
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; TITLE OF INVENTION: BIOLUMINESCENT NOVELTY ITEMS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
```



COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/911,020  
FILING DATE: 13-AUG-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/341,568  
FILING DATE: 22-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr, Gerald M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 365-262P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 536 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Trichoderma reesei  
STRAIN: QM9414  
US-08-911-020-1

Query Match 3.2%; Score 47; DB 2; Length 536;  
Best Local Similarity 64.0%; Pred. No. 0.00057;  
Matches 71; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 1364 aacaggctaataatggtttgtatctgaacatcgtgccatcgtaacatgaagcc 1423  
DB 411 AACCTTCTTACATGATGACTGCTATGACAAAGGTGTGGGTAGGTAAAGAAAAA 470  
QY 1424 tatatagcgggcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1474  
DB 471 AA 521

RESULT 13  
US-09-027-137-2  
Sequence 2, Application US/09027137  
Patent No. 6013450  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
APPLICANT: Yue, Henry  
TITLE OF INVENTION: CAF1-RELATED PROTEIN  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/027,137  
FILING DATE: Filed Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:

APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0476 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2852 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSNOT16  
CLONE: 2229466  
US-09-027-137-2

Query Match 3.2%; Score 47; DB 3; Length 2852;  
Best Local Similarity 59.3%; Pred. No. 0.0014;  
Matches 80; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1340 ttaagagagaagtaataacacagacagcctaatatagtggtttgtatcgaacatcg 1399  
DB 2432 TTAAGATGCTGTGTAATCTATTTTAAATTTTGTAAATATACATTTCTTTTA 2491  
QY 1400 gccatcgtacatcagtaagcctaatagcgggcaaaaaaaaaaaaaaaaaaaaaa 1459  
DB 2492 ACCACTGCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2551  
QY 1460 aaaaaaaaaaaaaa 1474  
DB 2552 AAAAAAAAAAAAAA 2566

RESULT 14  
US-08-897-340-1  
Sequence 1, Application US/08897340  
Patent No. 5955306  
GENERAL INFORMATION:  
APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.  
TITLE OF INVENTION: Weight Control Pathway Genes and Uses  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/897,340  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/715,032  
FILING DATE: 17-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Silverl, Jean M.  
REGISTRATION NUMBER: 39,030  
REFERENCE/DOCKET NUMBER: MNI-005CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941



Tue Oct 2 11:16:25 2001

us-09-537-654-1.rni

---

Page 9



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: October 1, 2001, 16:27:27 ; Search time 20.7 Seconds  
(without alignments)  
861.036 Million cell updates/sec

Title: US-09-537-654-2

Perfect score: 1530  
Sequence: 1 MGDQSGSNNGPOKXVYSGAQ.....VTGKGRDAVSSNNKRAVTV 294

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

A\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1530	100.0	294	22	AAV71974
2	1530	100.0	294	22	AAV71976
3	1450	94.8	281	22	AAV71975
4	576.5	37.7	376	21	AAV73340
5	324	21.2	350	21	AAV91937
6	323	21.1	350	19	AAW56264
7	323	21.1	350	21	AAV91932
8	313	20.5	350	19	AAW56265
9	313	20.5	350	21	AAV91933
10	300	19.6	286	21	AAW43527
11	300	19.6	344	21	AAW43526

12	300	19.6	372	21	AAW43525
13	299	19.5	344	19	AAW63042
14	283.5	18.5	340	20	AAW42376
15	281.5	18.4	339	16	AAW78183
16	281.5	18.4	339	19	AAW62523
17	281.5	18.4	339	21	AAW87918
18	280.5	18.3	339	15	AAW54070
19	280.5	18.3	339	16	AAW87181
20	280.5	18.3	339	19	AAW62522
21	280.5	18.3	339	21	AAW87917
22	279.5	18.3	343	22	AAV72088
23	278.5	18.2	344	22	AAV72086
24	276	18.0	365	16	AAW78182
25	271.5	17.7	340	20	AAW42375
26	243	15.9	346	21	AAW87243
27	200	13.1	189	21	AAW44428
28	153	10.0	159	21	AAW44429
29	137.5	9.0	128	21	AAW44430
30	126	8.2	781	17	AAW01871
31	113.5	7.4	461	20	AAV34463
32	113.5	7.4	462	20	AAV34339
33	107	7.0	451	13	AAW29636
34	104	6.8	386	21	AAW16691
35	103	6.7	386	20	AAV37531
36	101	6.6	273	16	AAW79145
37	101	6.6	273	19	AAW44140
38	101	6.6	347	19	AAW44138
39	97.5	6.4	198	21	AAW28750
40	97.5	6.4	384	21	AAW44386
41	96	6.3	438	18	AAW34668
42	95	6.2	763	21	AAW81751
43	95	6.2	764	21	AAW81563
44	94	6.1	487	20	AAV36888
45	93.5	6.1	340	17	AAW93198

## ALIGNMENTS

RESULT 1  
AAV71974  
ID AAV71974 standard; Protein; 294 AA.  
XX  
AC AAV71974;  
XX  
DT 28-MAR-2001 (first entry)  
XX  
DE Maize RAD51 orthologue protein #1.  
XX  
KW Maize; RAD51 orthologue; RAD52 epistasis group; double strand break; DSB;  
KW homologous recombination; transformation; gene targeting;  
KW transgenic plant; DNA repair.  
XX  
OS Zea mays.  
XX  
PN WC200068370-A2.  
XX  
PD 16-NOV-2000.  
XX  
PF 05-APR-2000; 2000MO-US09010.  
XX  
PR 05-MAY-1999; 9905-0132582.  
XX  
PA (PION-) PIONEER HI-BRED INT INC.  
XX  
PI Mahajan PB, Shi J;  
XX WPI; 2001-016092/02.  
XX N-PSDB; AAD02118.  
XX  
PT New maize RAD51 polynucleotides, useful for e.g. controlling  
PT recombination or transformation efficiency in plants, or as probes or  
PT amplification primers for detecting, quantifying or isolating gene

PT transcripts -  
 XX  
 PS Claim 11; Page 69-70; 76pp; English.  
 CC  
 CC The present sequence is a maize RAD51 orthologue protein #1. RAD51 is one  
 CC of the genes of the RAD52 epistasis group that is involved in double  
 CC strand breaks (DSBs) repair by homologous recombination. Control of  
 CC homologous recombination by modulating RAD51 is used for improving  
 CC transformation and gene targeting in transgenic plants. The RAD51  
 CC nucleotide may be used as probes or amplification primers for detecting,  
 CC quantifying or isolating gene transcripts, in detecting deficiencies in  
 CC the mRNA level during screening for desired transgenic plants, for  
 CC detecting gene mutations, for monitoring upregulation of expression or  
 CC changes in enzyme activity, for detecting any number of allelic variants,  
 CC orthologues or paralogues of the gene, or for site directed mutagenesis  
 CC in eukaryotic cells. It may also be used for recombinant expression of  
 CC its encoded polypeptide, or for use as immunogen in preparing and/or  
 CC screening of antibodies, and in sense or antisense suppression of one or  
 CC more genes in a host cell, tissue or plant. The RAD51 proteins may be  
 CC used in assays to agonise or antagonise the enzyme function, or as  
 CC immunogens or antigens to obtain antibodies.  
 CC  
 XX Sequence 294 AA:  
 SO  
 Query Match 100.0%; Score 1530; DB 22; Length 294;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-157;  
 Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MGQSSSRNGPQOKYVYSGAONAMDFSDLSQKHITTSQGLNDLIGGTHCKEYETIGG 60  
 DB 1 mgdsgsrngpqkkyvsqagawmdsldsqkhlttgsqdlndllggthckeyetl 60  
 QY 61 VPGVGKTLQGLAIINVOIPEECGLGKAVYIDTEGSMFVERVYOIAECIRDLIEHP 120  
 DB 61 vpgvgktlqglainvqipvecgglgkavayidtegsfmvervyiaecirdllehp 120  
 QY 121 HSHKSSSVQKQLQPERFLADIIYFRICSTEQIAYINMEKFLREHKDVRITVIDSVTF 180  
 DB 121 hshksssvqkqlqperfladiyfricsteqiayinmekflrehkdvrividsvtf 180  
 QY 181 HFRQDFEDIALRTRVLSGLSKLTKAKTYNLAIVLLNVTYTKFTSGPOLTLALDSWS 240  
 DB 181 hfrqdfedialtrvlsqslkmlaktynlavllnvtytkftsgpoltaladsws 240  
 QY 241 HSCNRLILHMNGENERYAHLDKSPSLPVASAPYAVTGKIRDAVSSNHKRAVY 294  
 DB 241 hscnrlilhmngeneryahldkspslpvasapyavtgkirdavssnhkravt 294  
 RESULT 2  
 ID AAY71976 standard; Protein; 294 AA.  
 XX  
 AC AAY71976;  
 XX  
 DT 28-MAR-2001 (first entry)  
 XX  
 DE Maize RAD51 orthologue protein #3.  
 XX  
 KW Maize; RAD51 orthologue; RAD52 epistasis group; double strand break; DSB;  
 KW homologous recombination; transformation; gene targeting;  
 KW transgenic plant; DNA repair.  
 XX  
 OS Zea mays.  
 XX  
 PN WO200068370-A2.  
 XX  
 PD 16-NOV-2000.  
 XX  
 PF 05-APR-2000; 2000MO-US09010.  
 XX  
 PR 05-MAY-1999; 99US-0132582.

XX  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Mahajan PB, Shi J;  
 XX  
 DR WPI; 2001-016092/02.  
 DR  
 DR N-PSDB; RAD02120.  
 XX  
 PT New maize RAD51 polynucleotides, useful for e.g. controlling  
 PT recombination or transformation efficiency in plants, or as probes or  
 PT amplification primers for detecting, quantifying or isolating gene  
 PT transcripts -  
 XX  
 PS Claim 1; Page 74-75; 76pp; English.  
 CC  
 CC The present sequence is a maize RAD51 orthologue protein #3. RAD51 is one  
 CC of the genes of the RAD52 epistasis group that is involved in double  
 CC strand breaks (DSBs) repair by homologous recombination. Control of  
 CC homologous recombination by modulating RAD51 is used for improving  
 CC transformation and gene targeting in transgenic plants. The RAD51  
 CC nucleotide may be used as probes or amplification primers for detecting,  
 CC quantifying or isolating gene transcripts, in detecting deficiencies in  
 CC the mRNA level during screening for desired transgenic plants, for  
 CC detecting gene mutations, for monitoring upregulation of expression or  
 CC changes in enzyme activity, for detecting any number of allelic variants,  
 CC orthologues or paralogues of the gene, or for site directed mutagenesis  
 CC in eukaryotic cells. It may also be used for recombinant expression of  
 CC its encoded polypeptide, or for use as immunogen in preparing and/or  
 CC screening of antibodies, and in sense or antisense suppression of one or  
 CC more genes in a host cell, tissue or plant. The RAD51 proteins may be  
 CC used in assays to agonise or antagonise the enzyme function, or as  
 CC immunogens or antigens to obtain antibodies.  
 CC  
 XX Sequence 294 AA:  
 SO  
 Query Match 100.0%; Score 1530; DB 22; Length 294;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-157;  
 Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MGQSSSRNGPQOKYVYSGAONAMDFSDLSQKHITTSQGLNDLIGGTHCKEYETIGG 60  
 DB 1 mgdsgsrngpqkkyvsqagawmdsldsqkhlttgsqdlndllggthckeyetl 60  
 QY 61 VPGVGKTLQGLAIINVOIPEECGLGKAVYIDTEGSMFVERVYOIAECIRDLIEHP 120  
 DB 61 vpgvgktlqglainvqipvecgglgkavayidtegsfmvervyiaecirdllehp 120  
 QY 121 HSHKSSSVQKQLQPERFLADIIYFRICSTEQIAYINMEKFLREHKDVRITVIDSVTF 180  
 DB 121 hshksssvqkqlqperfladiyfricsteqiayinmekflrehkdvrividsvtf 180  
 QY 181 HFRQDFEDIALRTRVLSGLSKLTKAKTYNLAIVLLNVTYTKFTSGPOLTLALDSWS 240  
 DB 181 hfrqdfedialtrvlsqslkmlaktynlavllnvtytkftsgpoltaladsws 240  
 QY 241 HSCNRLILHMNGENERYAHLDKSPSLPVASAPYAVTGKIRDAVSSNHKRAVY 294  
 DB 241 hscnrlilhmngeneryahldkspslpvasapyavtgkirdavssnhkravt 294  
 RESULT 3  
 ID AAY71975 standard; Protein; 281 AA.  
 XX  
 AC AAY71975;  
 XX  
 DT 28-MAR-2001 (first entry)  
 XX  
 DE Maize RAD51 orthologue protein #2.  
 XX  
 KW Maize; RAD51 orthologue; RAD52 epistasis group; double strand break; DSB;  
 KW homologous recombination; transformation; gene targeting;

KW transgenic plant; DNA repair.  
 XX  
 OS Zea mays.  
 XX  
 PN WO200068370-A2.  
 XX  
 PD 16-NOV-2000.  
 XX  
 PF 05-APR-2000; 2000WO-US09010.  
 XX  
 PR 05-MAY-1999; 99US-0132582.  
 XX  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Mahajan PB, Shi J;  
 XX  
 DR WPI: 2001-016092/02.  
 DR N-PSDB: AAD02119.  
 XX  
 PT New maize RAD51 polynucleotides, useful for e.g. controlling  
 PT recombination or transformation efficiency in plants, or as probes or  
 PT amplification primers for detecting, quantifying or isolating gene  
 PT transcripts  
 XX  
 PS Claim 11; Page 72; 76pp; English.  
 XX  
 CC The present sequence is a maize RAD51 orthologue protein #2. RAD51 is one  
 CC of the genes of the RAD52 epistasis group that is involved in double  
 CC strand breaks (DSBs) repair by homologous recombination. Control of  
 CC homologous recombination by modulating RAD51 is used for improving  
 CC transformation and gene targeting in transgenic plants. The RAD51  
 CC nucleotide may be used as probes or amplification primers for detecting,  
 CC quantifying or isolating gene transcripts, in detecting deficiencies in  
 CC the mRNA level during screening for desired transgenic plants, for  
 CC detecting gene mutations, for monitoring upregulation of expression or  
 CC changes in enzyme activity, for detecting any number of allelic variants,  
 CC orthologues or paralogues of the gene, or for site directed mutagenesis  
 CC in eukaryotic cells. It may also be used for recombinant expression of  
 CC its encoded polypeptide, or for use as immunogen in preparing and/or  
 CC screening of antibodies, and in sense or antisense suppression of one or  
 CC more genes in a host cell, tissue or plant. The RAD51 proteins may be  
 CC used in assays to agonise or antagonise the enzyme function, or as  
 CC immunogens or antigens to obtain antibodies.  
 CC  
 CC Sequence 281 AA;  
 SQ  
 QY 1 MGDDSGSRNGFQOKYVSAONAMDFSELSOKHITGSGDLNDILGGGIHCKREVTETIG 60  
 Db 1 mgddsgsrngfqpqkyvsagawamidselsqkhittsgdlnil199gihckevete19g 60  
 QY 61 VPVGKTKQLGQLAINVOIPECCGGLGKAVYIDTEGSSFWERYVQTAECGIRDIIEHP 120  
 Db 61 vpygkktqlgqlainvoipeccgglgkavyidtegsfwerwyvqtaecgirdiiehp 120  
 QY 121 HSHKSSSVKQLOLPERFLADIIYFRICSTTEQIAVINIMKEFLREHKDVRIVIDSVTF 180  
 Db 119 hshksssvkqlperfladiiyfricstteqiavinimkeflrehkdvriividsvtf 180  
 QY 181 HFRDDEFLARTRVLSGLSLKMKIARTYLAVALINOVTTKTEGSPOLLTALGDSWS 240  
 Db 179 hfrddeflartrvlsqslslkmiaktylavallnqvltktesqqltalagdsws 238  
 QY 241 HSCNTRLILHMNGERYAHLDKSPSPVASAPYAVTGKGIARD 282  
 Db 239 hscntrllilmngerneryahlkspslpvasapyavtgkgiard 280  
 RESULT 4

AAV73340  
 ID AAV73340 standard; Protein; 376 AA.  
 XX  
 AC AAV73340;  
 XX  
 DT 24-FEB-2000 (first entry)  
 XX  
 DE HTRM clone 2398682 protein sequence.  
 XX  
 KW HTRM, human transcriptional regulatory molecule; arteriosclerosis; AIDS;  
 KW arteriosclerosis; cirrhosis; cancer; leukaemia; diabetes mellitus;  
 KW Addison's disease; multiple sclerosis; rheumatoid arthritis; infection;  
 KW trauma; myasthenia gravis; adenocarcinoma; immune disorder; treatment.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9957144-A2.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 04-MAY-1999; 99WO-US09935.  
 XX  
 PR 05-MAY-1998; 98US-0084254.  
 PR 07-AUG-1998; 98US-0095827.  
 PR 02-OCT-1998; 98US-0102745.  
 XX  
 PA (INCY-) INCYTE PHARM INC.  
 XX  
 PI Hillman JL, Bandman O, Lal P, Yue H, Reddy R, Tang YT;  
 PI Gerstlin EH, Patterson C, Baughn MR, Azimzal Y, Lu DM;  
 DR WPI: 2000-052941/04.  
 DR N-PSDB: AA52425.  
 XX  
 PT New peptides useful for diagnosis, prevention and treatment of cancer  
 PT and immune disorders  
 XX  
 CC Claim 1; Page 100-101; 193pp; English.  
 XX  
 CC AAV73325-V73389 are human transcriptional regulator molecule (HTRM)  
 CC protein sequences. The HTRM protein and nucleotide sequences are useful  
 CC for preventing or treating disorders associated with decreased expression  
 CC or activity of HTRM which include cell proliferative disorders such as  
 CC arteriosclerosis and cirrhosis; cancers including adenocarcinoma and  
 CC leukaemia; immune disorders such as AIDS, Addison's disease, diabetes  
 CC mellitus, rheumatoid arthritis, multiple sclerosis, systemic lupus  
 CC erythematosus, and myasthenia gravis; infections and trauma. Antagonists  
 CC of the HTRM polypeptides are useful for treating or preventing disorders  
 CC associated with increased expression or activity of HTRMs. HTRM  
 CC polypeptides, their immunogenic fragments or oligopeptides are useful  
 CC for screening libraries of compounds in drug screening techniques.  
 CC Polynucleotides encoding HTRM are useful for blocking the transcription  
 CC of mRNA and regulating gene function by modulating the activity of  
 CC HTRM. Vectors expressing HTRM or agonists can also be used to prevent or  
 CC treat disorder associated with decreased HTRM expression. Antibodies  
 CC which specifically bind HTRM and polynucleotides encoding HTRM are useful  
 CC for diagnosing disorders associated with the expression of HTRM,  
 CC particularly in assays that detect the expression of HTRM. Nucleotide  
 CC sequences encoding HTRM may be useful to generate hybridization probes  
 CC useful in mapping the naturally occurring genomic sequence and to detect  
 CC differences in gene sequences among normal, carrier and affected  
 CC individuals. Using diagnostic assays, cancer can be detected prior to  
 CC the appearance of clinical symptoms and thereby progression of cancer can  
 CC be prevented by aggressive treatment or preventive measures.  
 CC  
 CC Sequence 376 AA;  
 SQ  
 QY Query Match 37.7%; Score 576.5; DB 21; Length 376;  
 Best Local Similarity 43.9%; Pred. No. 7.4e-54;  
 Matches 125; Conservative 47; Mismatches 98; Indels 15; Gaps 4;  
 QY 22 AWMDFSELSOKHITGSGDLNDILGGGIHCKREVTETIGVPGVKTQLGQLAINVOI 81

DB 87 alelleqetqgfilftcscaldilgygyplmkteicagpygkqlcmqldavtdqdp- 145  
 QY 82 EC-GGLGKAVYIDTEGSPWERYOIAEGCIRD-LEHFPHSHEKSSSVQKOLPERFL 139  
 DB 146 ecfgyagaeavfdtqgslmvdtrvdlatacqhqlkqkqgehrkalefdtdnll 205  
 QY 140 ADIYFRICSYTEQIAVINMKEFLREHKDVRIVIIDSVTFPHRODFELALRTYLSGL 199  
 DB 206 shlyyfticdytelqlqvyllpdlfshskvrlvldglafrmdldsltrllngl 265  
 QY 200 SLKLMRIAKTYNLAIVLNQVTKFTGSGFOLTLALGDSWSSCTNRLILHNGNERVAH 259  
 DB 266 agqmislammhravlltqmtkldngalllvpalgswgnaafrllfhwdrkqrilat 325  
 QY 260 LDKSPSLPVASAPYAVTGKIRDAV-----SSNHRAR 292  
 DB 326 lypspqkectvlfqlkpgqfrdtvtsacsllqtegsllstrkrxr 370

RESULT 5  
 AAY91937  
 ID AAY91937 standard; Protein; 350 AA.  
 AC AAY91937;  
 DT 19-JUL-2000 (first entry)  
 DE Ala163 modified human Rec2 serine kinase.  
 KW hsrRec2; serine kinase; homologous pairing; strand transfer; RAD51;  
 KW phosphorylation; cyclin E; p53; cell cycle.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Misc-difference 163 /note= "Y163A substitution from wild type"  
 FT MO200017329-A1.  
 PN 30-MAR-2000.  
 PD 17-SEP-1999; 99WO-US21642.  
 PF 21-SEP-1998; 98US-0157603.  
 PR (UYJR-) UNIV JEFFERSON THOMAS.  
 PA (CORR) CORNELL RES FOUND INC.  
 PA (KIME-) KIMERAGEN INC.  
 PI Havre PA, Rice MC, Holloman WK, Kmiec EB;  
 PI MPI; 2000-283562/24.  
 PT phosphorylating a serine-containing substrate by incubating it with  
 PT adenosine triphosphate and Rec2 kinase and measuring the level of  
 PT phosphorylation, useful for discovering specific antagonists or  
 PT agonists of Rec2  
 PS Claim 18; Page -; 41pp; English.

CC The present sequence is Ala163 hsrRec2, a human serine kinase modified by  
 CC a Y163A substitution. The hsrRec2 is in the same supergene family as the  
 CC mammalian protein having homologous pairing and strand transfer  
 CC activities, RAD51 and was isolated because of its homology to the  
 CC homologous pairing and strand transfer protein of *Ustilago maydis*. In  
 CC particular, hsrRec2 phosphorylates several proteins that control the cell  
 CC cycle, especially cyclin E and p53. The invention permits the  
 CC phosphorylation of cell cycle control proteins at sites that are  
 CC physiologically relevant. The invention can be practiced with either  
 CC murine or human Rec2 or a mutin or chimera of these proteins. In  
 CC particular the mutin has the sequence of a Rec2 kinase containing other

CC than a Tyr at position 163. The invention comprises a method of  
 CC phosphorylating a serine-containing substrate comprising incubating the  
 CC substrate with ATP (adenosine triphosphate) and Rec2 kinase or a  
 CC mammalian Rec2 and measuring the level of phosphorylation. The method is  
 CC useful for discovering compounds which are specific antagonists or  
 CC agonists of Rec2.  
 CC Note: The present sequence is not shown in the specification but is  
 CC derived from the hsrRec2 sequence given on pages 33-34  
 CC  
 CC Sequence 350 AA:  
 SO  
 Query Match 21.2%; Score 324; DB 21; Length 350;  
 Best Local Similarity 30.9%; Pred. No. 1.4e-26;  
 Matches 88; Conservative 50; Mismatches 99; Indels 48; Gaps 7;  
 QY 27 SDELSQKHITTGSGDLDNDLGGGHCKEVETGCGVPGVKTQLGOLAINVOLPVEGGL 86  
 DB 75 sadfspaflstlslaldealngvacgslteltpgpcgkqfclmmlslatlpnmgl 134  
 QY 87 GSKAVYIDTEGSPWERYOIAEGCIRDLEHFPHSHEKSSSVQKOLPERFL---ADIV 143  
 DB 135 egavvyldtesafsaerlveiae-----srprafnte-----ekllltskvh 178  
 QY 144 YFRICSYTEQIAVINMKEFLREHKDVRIVIIDSVTFHRODFE----DLALFRVLSG 198  
 DB 179 lyreltcddevlqrlsleeeils-kjklvllsdvasvvtkefdaglgqlkernkflar 237  
 QY 199 LSKLMRIAKTYNLAIVLNQVTKF-----TGSGFOLTLALG 236  
 DB 238 easslkylaefstlplvltnglthlsgalasqadlvsppaddslssegtsysscvcvlaig 297  
 QY 237 DSWSHSCTNRLILHNGNERVAHL-DKSPSLPVASAPYAVTGKIR 280  
 DB 298 ntwshtvntlllqyldserqlllaksplapstfsfytllkegl 342

RESULT 6  
 AAM56264  
 ID AAM56264 standard; Protein; 350 AA.  
 AC AAM56264;  
 DT 28-AUG-1998 (first entry)  
 DE Human recombinase hsrRec2 protein.  
 KW Recombinase; REC2; hsrRec2; homologous recombination; Irradiation;  
 KW apoptosis; cancer; human.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Region 4..7 /label= NLS  
 FT Region /note= "nuclear localisation sequence"  
 FT Region 108..116 /label= A-box  
 FT Region /note= "A box motif sequence"  
 FT Region 160..163 /label= B-box  
 FT Region /note= "Src-type phosphorylation site"  
 FT Region 186..199 /label= DNA  
 FT Region /note= "DNA binding sequence"  
 FT Region 206..210 /label= B-Box  
 FT Region /note= "B box motif"  
 PN W09811214-A1.  
 PD 19-MAR-1998.  
 PF 11-SEP-1997; 97WO-IB01217.





CC mammalian Rec2 and measuring the level of phosphorylation. The method is  
CC useful for discovering compounds which are specific antagonists or  
CC agonists of Rec2.

50 Sequence 350 AA;

Query Match	20.5%;	Score 313;	DB 21;	Length 350;
Best Local Similarity	32.1%;	Pred. No. 2.2e-25;		
Matches 100;	Conservative 46;	Mismatches 98;	Indels 68;	Gaps 13;

```

QY 20 QANMDD---FSELSQKHIITSSDLDNDIGGJHCKEWTEJAGVPGVKGKJGLOIAIN 76
Db 65 qtaeyelktrrhshspafstlclaldeaahgvyprgssteltppgpgcrgqfclmmsv 124
QY 77 VOJRVCCGLGSKANYIDTESFMEVENVYDIAEGCJRDILHEHP---SHSEK---SSSV 129
Db 125 atlpsrlgglegavuyldltesafaelvae-----srfpyftheekl1lltsrv 177
QY 130 Q--KOLQPRFLADIVYFRICSTYEQJAVINVMKEFLREHKVKVRIVIDSYTFHPRDOFE 187
Db 178 hlacrltceglq-----rlsleeeell-----skvklivdsiasvrlfejd 221
QY 188 -----DLAIFRVL--SGSLKLMKIATYMLANVLLNOYTTKF-----224
Db 222 pklqglkierknflfgkgsal--lkylaeafisipvltnqatlhtlsqalpsqadlvsyddi 280
QY 225 -----TEGSFOKLTLGSMGSHSCSTNLLHMMNGENERYAL--DKSPSLPVASAPVATYCK 278
Db 281 slssegysgscvalaagltgwtbcvntclllqydsertqllakaplaatsfyytlkge 340
QY 279 GIRDANSSNHR 290
Db 341 gl---vlgqher 349

```

## RESULT 10

ID AAG43527 standard; Protein; 286 AA.

AC AAG43527;

DT	18-OCT-2000	(first entry)
XY		

Zea mays protein fragment SEQ ID NO: 54415

KW protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.

**Zea mays subsp. ma**

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.  
VV

PR	25-FEB-1999	9905-0121825
PR	05-MAR-1999	9905-0121380
PR	09-MAR-1999	9905-0123548
PR	23-MAR-1999	9905-0125788
PR	25-MAR-1999	9905-0126564
PR	29-MAR-1999	9905-0126785
PR	01-APR-1999	9905-0127462
PR	06-APR-1999	9905-0128234
PR	08-APR-1999	9905-0128714
PR	16-APR-1999	9905-0129845
PR	19-APR-1999	9905-0130077
PR	21-APR-1999	9905-0130449
PR	23-APR-1999	9905-0130591
PR	28-APR-1999	9905-0131449

PR	30-APR-1999	9905-0132048
PR	30-APR-1999	9905-0132048
PR	04-MAY-1999	9905-0132484
PR	05-MAY-1999	9905-0132485
PR	06-MAY-1999	9905-0132486
PR	07-MAY-1999	9905-0132487
PR	11-MAY-1999	9905-0134256
PR	14-MAY-1999	9905-0134218
PR	14-MAY-1999	9905-0134219
PR	14-MAY-1999	9905-0134221
PR	18-MAY-1999	9905-0134768
PR	19-MAY-1999	9905-0134941
PR	20-MAY-1999	9905-0135124
PR	21-MAY-1999	9905-0135553
PR	24-MAY-1999	9905-0135629
PR	25-MAY-1999	9905-0136021
PR	27-MAY-1999	9905-0136392
PR	28-MAY-1999	9905-0136782
PR	01-JUN-1999	9905-0137222
PR	03-JUN-1999	9905-0137528
PR	04-JUN-1999	9905-0137502
PR	07-JUN-1999	9905-0137724
PR	08-JUN-1999	9905-0138040
PR	10-JUN-1999	9905-0138549
PR	10-JUN-1999	9905-0138847
PR	14-JUN-1999	9905-0139119
PR	16-JUN-1999	9905-0139452
PR	16-JUN-1999	9905-0139453
PR	17-JUN-1999	9905-0139492
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139455
PR	18-JUN-1999	9905-0139456
PR	18-JUN-1999	9905-0139457
PR	18-JUN-1999	9905-0139458
PR	18-JUN-1999	9905-0139459
PR	18-JUN-1999	9905-0139460
PR	18-JUN-1999	9905-0139461
PR	18-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0140353
PR	23-JUN-1999	9905-0140354
PR	24-JUN-1999	9905-0140695
PR	28-JUN-1999	9905-0140823
PR	29-JUN-1999	9905-0140991
PR	30-JUN-1999	9905-0141287
PR	01-JUL-1999	9905-0141287
PR	01-JUL-1999	9905-0141287
PR	02-JUL-1999	9905-0142154
PR	06-JUL-1999	9905-0142051
PR	06-JUL-1999	9905-0142390
PR	09-JUL-1999	9905-0142820
PR	12-JUL-1999	9905-0142877
PR	13-JUL-1999	9905-0142877
PR	14-JUL-1999	9905-0143342
PR	15-JUL-1999	9905-0143624
PR	16-JUL-1999	9905-0144005
PR	16-JUL-1999	9905-0144086
PR	19-JUL-1999	9905-0144325
PR	19-JUL-1999	9905-0144331
PR	19-JUL-1999	9905-0144332
PR	19-JUL-1999	9905-0144333
PR	19-JUL-1999	9905-0144334
PR	19-JUL-1999	9905-0144335
PR	20-JUL-1999	9905-0144632
PR	20-JUL-1999	9905-0144684
PR	21-JUL-1999	9905-0144814



PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139455.  
PR 16-JUN-1999; 99US-0139455.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139765.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151308.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154079.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.

PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160811.  
PR 22-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 26-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 28-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 19.6%; Score 300; DB 21; Length 344;  
Best Local Similarity 33.1%; Pred. No. 5,5e-24;

Matches 88; Conservative 41; Mismatches 109; Indels 28; Gaps 5;

OY 24 DMFSELQKHTTGGSLNDILGSGHCKEVTETIGVGVGKQGLQALINVOIPEEC 83  
DB 97 dIlLrkKsvnlItgsgalndelIggletlcItaefgfsrgtqlahltlcvtqlpIhm 156  
OY 84 GILGKAVYIDPESFMYRYOIAEGCIRDLLEHFRSHKSSVQKOLPEFTADY 143  
DB 157 hgngskvayidtegtffrperlvplae-----ffgmदानavlhl 196  
OY 144 YERICSYTEQIAYINMEKFLREHKDVRIVIIDSVTFHRODPE---DLALRTFVLSGLS 200  
DB 197 YarelyehgynllIglakmae-epfklllvdsvalfrvdfsgrelaerqkIagml 255  
OY 201 LKMKIKATYLAVALNQVTTKTESGFS---QUTLALGDSWSHSCNRLILH-WNGNER 256  
DB 256 srlkIaeetvayltnqyIadpbggmIltcdpkkpagnhlaaetlrlmltkgqegr 315  
OY 257 YAHLDKSPSLPVASAPYAVTGKGRD 282  
DB 316 vckIfdnpIpegeavfvtvsggImd 341

RESULT 12

ID AAG43525 standard; Protein; 372 AA.

AC AAG43525;

DT 18-OCT-2000 (first entry)

DE Zea mays protein fragment SEQ ID NO: 54413.

KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence; corn.

OS Zea mays subsp. mays.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127452.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134421.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.



KW Arabidopsis thaliana meiosis-specific DMCI gene; AtDMCI1 promoter;  
 KW meiosis specific expression; ablation; meiotic cell; isolation;  
 KW apomictic plant; increase; meiotic recombination; introgression;  
 KW sterile plant; seed production.

OS Arabidopsis thaliana.

PN W09828431-A1.

PD 02-JUL-1998.

PF 24-DEC-1997; 97WO-GB03546.

PR 24-DEC-1996; 96GB-0026858.

PA (INNE-) INNES CENT INNOVATIONS LTD JOHN.

PI Dicks R, Jones JDG, Klimyuk VI;

DR WPI; 1998-377661/32.

PT New isolated Arabidopsis meiosis-specific promoter - useful for  
 PT meiosis-specific transcription of genes, e.g. for isolation of  
 PT apomictic plants or removing DNA from transgenic plants

PS Disclosure; Fig 7; 69pp; English.

CC The present sequence is encoded by the Arabidopsis thaliana  
 CC meiosis-specific DMCI gene (AtDMCI1). The AtDMCI1 promoter and homologues  
 CC can be used to confer meiosis specific expression on a sequence operably  
 CC linked to the promoter. They can be used for e.g. ablation of meiotic  
 CC cells and isolation of apomictic plants, designing an efficient  
 CC homologous recombination system for plants, increasing meiotic  
 CC recombination frequency, for introgression of alien chromosome segments  
 CC into host plant, or altering normal events of cell cycle during the time  
 CC of meiosis and producing male and female sterile plants. The promoters  
 CC can also be used in searching for apomictic mutants or used by seed  
 CC producers to produce seeds apomictically. They can also be used for  
 CC removing any unwanted DNA sequences from transgenic plants.

SQ Sequence 344 AA;

Query Match 19.5%; Score 299; DB 19; Length 344;  
 Best Local Similarity 32.6%; Pred. No. 7e-24;

Matches 90; Conservative 42; Mismatches 112; Indels 32; Gaps 6;

OY 15 YVSGAQNAMDMSDELSOKHTTGGDNDILGCGHCKEVEYIGVPGVGTOLGLOLA 74

DB 92 ymgs-----dalikrksvvtctgcqaldlll99gletsatleaefrsgtqlahlc 147

OY 75 INVQIPVECGGLGKRAVYIDTEGSPMERVYQIAGCIRDLILHPPHSHKSSSVOKOLO 134

DB 148 vctqptlmkkgngkvayidtegtftrpdiylpiae-----rfgmd 187

OY 135 PERFLADYIFRRCISTYBOIAVINMEKFLREHKDVRIVIDSYTFHRDPE---DLAL 191

DB 188 pvaivdnlilyataytebqyalllglakmsee-epfrillvdsiallfrvfttgrgelad 246

OY 192 FRRVIVSGLSLKLAKTYNLAVALVNLNOYTKFTBGSF---QLTLALGDSMSHSCNLI 248

DB 247 rsgkqlagmrlrliklaeevnavymtnlvgadprrgmflsdpkkrpagnviahnaatlill 306

OY 249 LH-WNGNERYAHLDKSPSLPVASAPYAVTGKIRDA 283

DB 307 frkgkydtrvcvkvydaplnaeeasfgiltggyiada 342

RESULT 14

AA142376

ID AA142376 standard; Protein: 340 AA.

XX AC AA142376;

XX 10-DEC-1999 (first entry)

DT ZMRAD51B amino acid sequence.

DE recombinase; gene targeting; gene insertion; transgenic plant;

KW expression vector; plasmid; meiosis-specific homologue.

XX Zea mays.

OS W03941394-A1.

PN 19-AUG-1999.

PD 11-FEB-1999; 99WO-US02900.

PF 13-FEB-1998; 98US-0074745.

PR (PION-) PIONEER HI-BRED INT INC.

PA Bowen BA, Chamberlin MA, Drummond BJ, McElver JA;

PI Rothstein RJ;

DR WPI; 1999-527374/44.

DR N-PADB; AA211870.

PT Maize recombinase genes ZMRAD51A and ZMRAD51B useful for gene  
 PT targeting system and production of transgenic plants

PS Claim 8; Page 42; 66pp; English.

CC This is the amino acid sequence of the ZMRAD51B recombinase enzyme.  
 CC The RAD51 recombinase from *S.cerevisiae* has a role in the catalysis  
 CC of the in vitro pairing and strand exchange between circular viral  
 CC single strand DNA and linear duplex DNA. The identification of the  
 CC ZMRAD51 genes will allow the identification and characterisation of  
 CC the functional activities of recombinase enzymes from plants.  
 CC The two isolated cDNAs (AA211868 and AA211872) are more closely related  
 CC to RAD51 family members expressed in mitotic cells than to the  
 CC meiosis-specific homologues from plants (LIM15) and yeast (DMC1).  
 CC Despite the significant transcription levels of the genes, that have  
 CC been detected primarily in immature ears and anthers that contain cells  
 CC progressing through the early stages of meiosis.  
 CC The sequences claimed by the invention can be used to modulate the  
 CC activity of RAD51 recombinase in transgenic plants, especially in maize.  
 CC The genes can be used in gene targeting systems and for generating  
 CC transgenic plants, especially in agriculturally important crop plants.

SQ Sequence 340 AA;

Query Match 18.5%; Score 283.5; DB 20; Length 340;  
 Best Local Similarity 29.9%; Pred. No. 3.3e-22;

Matches 83; Conservative 50; Mismatches 106; Indels 39; Gaps 7;

OY 18 GAONAMFMSDELSOKHTTGGDNDILGCGHCKEVEYIGVPGVGTOLGLOLAIV 77

DB 86 gftsaasqihqatlelqiltgtsrelldgltetstemyefsgtkqlchtlcvrc 145

OY 78 QIPECGGLGKRAVYIDTEGSPMERVYQIAE-----GCRDLILHPPHSHKSSSVOKO 132

DB 146 qipldgggggkalyidaegtfprqllqdadfngina--dvlenavayaraynlthgsr 203

OY 133 LQPERFLADYIFRRCISTYBOIAVINMEKFLREHKDVRIVIDSYTFHRDPE---DL 189

DB 204 lile-----aasmvretf-----almvdsatalyrtidfsgrgl 239

OY 190 ALRFRVIVSGLSLKLAKTYNLAVALVNLNOYTKFTBGSF---QLTLALGDSMSHSCN 245

DB 240 sarqmbhlakflrslqkladeqyavvnlngvavqdamfagpikpignlmahast 299

OY 246 RLILH-WNGNERYAHLDKSPSLPVASAPYAVTGKIRD 282

Db 300 rlfirkygeerickvispciaaeafqissegvtd 337

## RESULT 15

ID AAR78183 standard; Protein; 339 AA.

AC AAR78183;

DT 15-FEB-1996 (first entry)

DE Human RAD51 protein, involved in DNA repair.

KW DNA mismatch repair; DNA recombination; site-specific; unwinding;

KW right-handed helical nucleoprotein; RAD51; human; recA;

topoisomerase.

OS Homo sapiens.

PN JP07143890-A.

PD 06-JUN-1995.

PF 28-MAY-1993; 93JP-0127594.

PR 28-MAY-1993; 93JP-0127594.

PA (TOYM ) TOYBO KK.

DR WPI: 1995-236467/31.

DR N-PSDB; AAQ91763.

PT Structural gene encoding RAD51, used in production of RAD51 - used to study drugs against diseases caused by DNA damage, e.g. by UV or X radiation, and to improve efficiency of gene therapy targeting

PS Claim 1; Page 16-17; 19pp; Japanese.

CC DNA coding for a RAD51 protein having the amino acid sequence in AAR78183 is claimed. The RAD51 protein, which contains an ATP-binding domain, is involved in DNA repair of mismatched base pairs and site-specific recombination in antibody gene rearrangements. The protein binds to double-stranded DNA to form a right-handed helical nucleoprotein that extends (by 1.5 times) the pitch of B-form DNA. As a result, the helix is unwound. The RAD51 protein will be useful in designing drugs to treat diseases associated with environmental DNA damage.

CC Sequence 339 AA;

Query Match 18.4%; Score 281.5; DB 16; Length 339;

Best Local Similarity 30.4%; Pred. No. 5.4e-22;

Matches 80; Conservative 46; Mismatches 96; Indels 41; Gaps 7;

QY 35 ITTSGDLNDLGGIGCKEYETIGVPGVGTOLGIOLAINVOIPVCGGLGKAVYID 94

DB 102 ltskseldkllgylegsltemfgrtqqlchtltavtcqjldirgsggkamyid 161

QY 95 TEGSFNWERVOYIAE-----GCIRDLEHFPHSHEKSSVOKOLQPERFLADITYFRICS 149

DB 162 tegtfreperllavaerylg--sdvldnvayaraftdhtqql----- 204

QY 150 YTEQIAVINMEKFLREHKDRIYIDSVTFHFRDFE---DLALRTVLGSLKLMKI 206

DB 205 -----ygasamwvseyallivdsatalyrtidysgrgelsarqmhlarfmlrl 255

QY 207 AKTYNLAVLLNOVTKFTGSGFQLT-----ALGDSMSHCNRLILH-WNGNERAHL 260

DB 256 adelgavavltngvvaq--vdgaamfaadpkpignlianastrlylrkygetricki 314

QY 261 DKSPSLFVASAPYAVTGGIRDA 283

Db 315 ydspclpeaamfainadgvgda 337

Search completed: October 1, 2001, 16:28:15  
Job time: 48 sec

Tue Oct 2 11:16:27 2001

us-09-537-654-2.rag

---

Page 14

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 04:52:04 ; Search time 2002.11 Seconds  
(without alignments)  
11387.717 Million cell updates/sec

Title: US-09-537-654-1

Perfect score: 1474  
Sequence: 1 tcgaccacgcgcgcgcact.....aaaaaaaaaaaaaaaaaaaaa 1474

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_ba3: \*  
4: gb\_in1: \*  
5: gb\_in2: \*  
6: gb\_in3: \*  
7: gb\_om: \*  
8: gb\_ov: \*  
9: gb\_pat1: \*  
10: gb\_pat2: \*  
11: gb\_ph: \*  
12: gb\_pl1: \*  
13: gb\_pl2: \*  
14: gb\_pl3: \*  
15: gb\_pl4: \*  
16: em\_ba1: \*  
17: em\_ba2: \*  
18: em\_fun: \*  
19: em\_htgo\_hum: \*  
20: em\_htgo\_inv: \*  
21: em\_htgo\_rtd: \*  
22: em\_htg\_hum1: \*  
23: em\_htg\_hum2: \*  
24: em\_htg\_hum3: \*  
25: em\_htg\_hum4: \*  
26: em\_htg\_hum5: \*  
27: em\_htg\_hum6: \*  
28: em\_htg\_hum7: \*  
29: em\_htg\_hum8: \*  
30: em\_htg\_inv1: \*  
31: em\_htg\_inv2: \*  
32: em\_htg\_other: \*  
33: em\_htg\_rtd: \*  
34: em\_hum1: \*  
35: em\_hum2: \*  
36: em\_hum3: \*  
37: em\_hum4: \*  
38: em\_hum5: \*  
39: em\_hum6: \*  
40: em\_hum7: \*  
41: em\_in: \*  
42: em\_om: \*  
43: em\_or: \*

44: em\_ov: \*  
45: em\_pat: \*  
46: em\_ph: \*  
47: em\_pl: \*  
48: em\_ro: \*  
49: em\_rtd: \*  
50: em\_sy: \*  
51: em\_un: \*  
52: em\_v1: \*  
53: gb\_sts1: \*  
54: gb\_sts2: \*  
55: gb\_sts3: \*  
56: gb\_sy: \*  
57: gb\_un: \*  
58: gb\_v11: \*  
59: gb\_v12: \*  
60: gb\_hcg1: \*  
61: gb\_hcg2: \*  
62: gb\_hcg3: \*  
63: gb\_hcg4: \*  
64: gb\_hcg5: \*  
65: gb\_hcg6: \*  
66: gb\_hcg7: \*  
67: gb\_hcg8: \*  
68: gb\_hcg9: \*  
69: gb\_hcg10: \*  
70: gb\_hcg11: \*  
71: gb\_hcg12: \*  
72: gb\_hcg13: \*  
73: gb\_hcg14: \*  
74: gb\_hcg15: \*  
75: gb\_hcg16: \*  
76: gb\_hcg17: \*  
77: gb\_hcg18: \*  
78: gb\_hcg19: \*  
79: gb\_hcg20: \*  
80: gb\_hcg21: \*  
81: gb\_hcg22: \*  
82: gb\_hcg23: \*  
83: gb\_hcg24: \*  
84: gb\_hcg25: \*  
85: gb\_pr1: \*  
86: gb\_pr2: \*  
87: gb\_pr3: \*  
88: gb\_pr4: \*  
89: gb\_pr5: \*  
90: gb\_pr6: \*  
91: gb\_pr7: \*  
92: gb\_pr8: \*  
93: gb\_pr9: \*  
94: gb\_rtd: \*  
95: gb\_rtd2: \*  
96: gb\_in4: \*  
97: gb\_pr10: \*  
98: em\_ba3: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1474	100.0	1474	9	AX046601 Sequence
2	1305.4	88.6	1333	9	AX046605 Sequence
3	1243.2	84.3	1459	9	AX046603 Sequence
4	358.8	24.3	1101	13	ATN299424 Sequence
5	178.4	12.1	1295	88	AF029669 Homo sapi
6	171	11.6	13652	83	AP003310 Oryza sat
7	171	11.6	163419	13	AP002908 Oryza sat
8	103.4	7.0	1123	91	BC000667 Homo sapi

9	101.4	6.9	2340	13	ATH296174	AJ296174 Arabidops
10	101.4	6.9	122871	12	AC002387	AC002387 Arabidops
11	87.2	5.9	110000	84	LMFICR32_17	Continuation (18 o
12	77.8	5.3	10980	1	AE001035	AE001035 Archaeogl
13	74.6	5.1	13968	1	AE000901	AE000901 Methanoba
14	72	4.9	1699	3	SS045310	U4310 Sulfolobus
15	71.4	4.8	202000	2	AP000058	AP000058 Aeropyrum
16	70	4.7	1241	14	LEU22441	U22441 Lycopersico
17	68.6	4.7	1995	14	CAU39808	U39808 Candida alb
18	67.2	4.6	1740	14	DE4035	DE4035 Schizosacch
19	67.2	4.6	4230	12	AB008545	AB008545 Schizosac
20	67.2	4.6	15350	12	SPAC8E11	AL021817 S. pombe c
21	66.4	4.5	844	71	AC029940	AC029940 Giardia 1
22	66.4	4.5	992	75	AC074448	AC074448 Giardia 1
23	66.4	4.5	1030	72	AC058893	AC058893 Giardia 1
24	66.4	4.5	1050	75	AC078679	AC078679 Giardia 1
25	65.8	4.5	1551	12	AF079428	AF079428 Zea mays
26	65	4.4	1585	9	AX047126	AX047126 Sequence
27	64.4	4.4	1551	12	AF079429	AF079429 Zea mays
28	64.2	4.4	813	53	CNS075T2	AL430524 clone BAO
29	63.6	4.3	2745	5	AF198107	AF198107 Giardia 1
30	62.8	4.3	1177	14	ATU43528	U43528 Arabidopsis
31	62.8	4.3	1219	9	AX047122	AX047122 Sequence
32	62.6	4.2	1020	7	AF017729	AF017729 Oryctolag
33	62.4	4.2	1800	15	YSCDWCL1	YSCDWCL1 Saccharomyc
34	62.4	4.2	2774	14	SCE9163	DI0865 Yeast genes
35	62.4	4.2	64044	14	SCE9163	U18922 Saccharomyc
36	61.2	4.2	1518	1	AB026115	AB026115 Pyrobacul
37	60.6	4.1	921	72	AC066858	AC066858 Giardia 1
38	60.6	4.1	928	72	AC045789	AC045789 Giardia 1
39	60.2	4.1	1101	72	AC053814	AC053814 Giardia 1
40	59.8	4.1	1260	1	AF145465	AF145465 Desulfuro
41	59.4	4.0	1502	1	AF052597	AF052597 Pyrococcu
42	58.6	4.0	923	53	CNS060D1	AL410915 T3 end of
43	58.4	4.0	971	53	CNS060RED	AL411887 T7 end of
44	58.4	4.0	1021	53	CNS0606W	AL410286 T3 end of
45	58.4	4.0	147620	62	AC011195	AC011195 Homo sapi

## ALIGNMENTS

RESULT 1  
AX046601 1474 bp DNA PAT 15-DEC-2000  
LOCUS Sequence 1 from Patent WO0068370.  
DEFINITION AX046601  
ACCESSION AX046601  
VERSION AX046601.1 GI:11876173  
KEYWORDS  
SOURCE Zea mays.  
ORGANISM Zea mays.  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 1474)  
Mahaajan, P.B. and Shi, J.  
Rad51 from maize  
Patent: WO 0068370-A 1 16-NOV-2000;  
PIONEER HI-BRED INTERNATIONAL, INC. (US)  
Location/Qualifiers  
1..1474  
/organism="Zea mays"  
/db\_xref="taxon:4577"  
310..1194  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC18901.1"  
/db\_xref="GI:11876174"  
/translation="MGDSGRNGPQOKYVSGAQNAMDFSDLSQKHITGSGDLND  
ILGGIHCKEVTEIGGVGVKQGLQIAINVOIIPVEGSGGKAVYIDEGSPME  
RYVQIAEGCIRIDLEHPSHESKSSVVOLOPEFLADYIFRICSYTEQYINVM  
EKPIRHKVRIIVITDSVHFHRDOPEDIALPTRYLSGISIKIMKIATYINAVILN  
OVTTRTESFQITLALGDSMSHCTNRLILHMNGNERIAHLDKPSLPVASAPAVT

BASE COUNT	415 a	337 c	369 g	353 t
ORIGIN	GKIRDVASSNNKRAVTT			
Query Match	100.0%; Score 1474; DB 9; Length 1474;			
Best Local Similarity	100.0%; Pred. No. 0;			
Matches 1474; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
1	tcgagccacggttcgacactgactgactccacgtccacacgttgacagttgctgtcc	60		
1	TCGAGCCACGGTTCGACACTGACTCCACGTTCACGTGCGCAATTCGTTGGTCCC	60		
61	cggagcccaaaagcgcggtgagccgagcccgagacgagcgcgcgactccc	120		
61	CGGAGCCCAAAAGCGCGGTGAGCCGAGCCCGAGACGAGCGCGCGCACTCCC	120		
121	cctaagcgacgagcgcggtcgagcgtgtaagcggtggtggtggtggtggtggt	180		
121	CCTAAGCGACGAGCGGTCGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGT	180		
181	gagcggtgtgagcgagatagagagtgaggtgaggtgaggtgaggtgaggtgag	240		
181	GAGCGGTGTGAGCGAGATAGAGAGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG	240		
241	gatactgaatcgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgac	300		
241	GATACTGAATCGACGTGACGTGACGTGACGTGACGTGACGTGACGTGACGTGAC	300		
301	gcaatcacacgagagagatcaatcgtggtcctagaagatgagacacagagatc	360		
301	GCAATCACACGAGAGAGATCAATCGTGTCTCTAGAAATGAGACACAGAGATCT	360		
361	ggaagcccaagatgctcgtgagatgctcgtgagatgctcgtgagatgctcgtg	420		
361	GGAAGCCCAAGATGCTCGTGAGATGCTCGTGAGATGCTCGTGAGATGCTCGTG	420		
421	ggttcgtgagcctcctacatgacatggtggtggtggtggtggtggtggtggtg	480		
421	GTTCTGTGAGCCTCCTACATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	480		
481	atcggtggtcgtccagaggtgtgtgtaaacctcaactcgtggtggtggtggtg	540		
481	ATCGGTGTCGTCCAGAGGTGTGTGTAACCTCAACTCGTGTGTGTGTGTGTGTG	540		
541	caaatcccaagtgagatgtgtgtggtcgtggtggtggtggtggtggtggtggt	600		
541	CAATCCCAAGTGAGATGTGTGTGTCGTGTGTGTGTGTGTGTGTGTGTGTGTGT	600		
601	agtttcacgtgtgagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	660		
601	AGTTTCACGTGTGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	660		
661	caatttcgacacgacgagagatgctcctcgtgtccaaacacatcacagctgag	720		
661	CATTTCGACACGACGAGAGATGCTCCTCGTGTCCAAACACATCACAGCTGAG	720		
721	tttcctgagatcattactcctcagatgagatgagatgagatgagatgagatgag	780		
721	TTTCCTGAGATCATTACTCCTCAGATGAGATGAGATGAGATGAGATGAGATGAG	780		
781	aactcacatgagagatgctcctcagatgagatgagatgagatgagatgagatgag	840		
781	AACTCACATGAGAGATGCTCCTCAGATGAGATGAGATGAGATGAGATGAGATGAG	840		
841	gttacttcacattcgtcagacagatgttgagatcgtgacgtgagacagatgag	900		
841	GTTACTTCACATTCTGTCAGACAGATGTGAGATCGTGACGTGAGACAGATGAG	900		
901	ggatatacatgagatgagatgagatgagatgagatgagatgagatgagatgag	960		
901	GGATATACATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG	960		

QY 961 aaccaagtcactactaaattacagaaggtcattcaattgactctgtctagtgac 1020  
|||||  
Db 961 AACCAAGTCACTACTAAATTATACAGAGGTCATTCAATTGACTCTGTCTAGTGAC 1020  
QY 1021 agctggtccactcatatgacgaacgggttgattctgcactcgtgaatggaacgaatc 1080  
|||||  
Db 1021 AGCTGCTCCACTCATATGACGAACCGGTGTGATTCGTGACATGGAAAGGGAACGATAC 1080  
QY 1081 gcaactctgataagctccctcactcccaagtaagcccaagcgtatgcaatgacagc 1140  
|||||  
Db 1081 GCACATCTGTGATTAAGTCTCCTTCACTTCCAGTACGCTCAGACCGCTATGCAAGTACAGGC 1140  
QY 1141 aaaggatagaagatgctgtagctcaaacacacgaagcgaagccgaatgaatgacattc 1200  
|||||  
Db 1141 AAAGGATTAAGATGCTGTGAGCTCAACACAGGAGAGCCGAGTAAGTACGATTCCT 1200  
QY 1201 tgggtcaagacttgatgtatgtccactacgtctcctgcaagcttcttcgcaatgacattc 1260  
|||||  
Db 1201 TGGTGTCAAGCATTGTATGTCACACTACGCTCCTGACAGCTTCTTCCGATGATCTTTT 1260  
QY 1261 ggaactgaaggtgagactgagagataagtaacattcttgatctcaatgctgtctgac 1320  
|||||  
Db 1261 GGACTAGTAGAGTGAAGTGAAGTAAGTACCATTTGTTGATTTCTAGTTGCTTGTGTC 1320  
QY 1321 cgttggtaccacacacacttaagagaagaataacacagaacaggtcattatagtg 1380  
|||||  
Db 1321 CGTTGGCTACCAACCAACTTAAGAGAGAAGTAATACACAGAACAGGCTAATATATGTG 1380  
QY 1381 ttgtgtatcgaatctgtgcccactgtacatttgaagtaagccatataatgagcgcaaaa 1440  
|||||  
Db 1381 TTTGTATCTGAACATCTGCGCCATGCTACATTCATGAAGCCATATATATAGCGGCAAAA 1440  
QY 1441 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1474  
Db 1441 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1474

RESULT 2  
AX046605 1333 bp DNA PAT 15-DEC-2000  
LOCUS Sequence 5 from Patent WO0068370.  
DEFINITION AX046605  
ACCESSION AX046605  
VERSION AX046605.1 GI:11876177  
KEYWORDS  
SOURCE  
ORGANISM  
Zea mays.  
Zea mays.  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
Panicoideae; Andropogoneae; Zea.  
REFERENCE  
1 (bases 1 to 1333)  
AUTHORS Mahajan, P.B. and Shi, J.  
TITLE Rad51 from maize  
JOURNAL Patent: WO 0068370-A 5 16-NOV-2000;  
PIONEER HI-BRED INTERNATIONAL, INC. (US)  
FEATURES  
source  
1.1333  
Location/Qualifiers  
/organism="Zea mays"  
/db\_xref="taxon:4577"  
169..1053  
CDS  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC18903.1"  
/db\_xref="GI:11876178"  
translation="MGDSSGRNGPOKVVSGAONAMWDFSDLSOKHTTGGSLND  
ITGGTHCKEVEITGSGVGVKTOGLDIAIVQIPVEGCGAGKRAYIDTGSFME  
RYQIAEGCIRDLERFPHSHKSSSVQKQLODPERLADITFRICSTEQIAYINM  
EKFREHKDVRIVITDSVTFHRQDEDLALRRLVLSGLIKMLIAKTYNLAVALN  
QVTKPTGESFQITLALGDSWSHSCFNRLILHMNGERNERAHDKSPSLTPASAPYAVT  
GKGRDAVSNNHRAVYT"  
BASE COUNT 377 a 284 c 326 g 345 t 1 others  
ORIGIN

Query Match 88.6%; Score 1305.4; DB 9; Length 1333;  
Best Local Similarity 98.6%; Pred. No. 0;  
Matches 1315; Conservative 1; Mismatches 17; Indels 0; Gaps 0;  
QY 142 cgaagtaagcgcgtcgtgctgagccacgaagcaggtacagcgcgtgtgagagagata 201  
|||||  
Db 1 CGACGTAAAGCGGGTGGCTGAGGCGCCGACGAGGAGGTACGACGGTGTGGAGGAGATA 60  
QY 202 ttggaaggtgaggtgactcaacgggttgaggcgtgtgatactgaatccgcactga 261  
|||||  
Db 61 TGAGAGGTGAGGTGCTCAACGGGTGGGCTGTGTGATCTGAATCCGCACTGTA 120  
QY 262 gtctctcttcccccacatcagtaacacctcccaagtgtgcaatcaccaatggagataca 321  
|||||  
Db 121 GTTCTCTTCTTCCCCCAATACAGTACCACTTCCAAAGTGGCAATACATGGGAGATCA 180  
QY 322 tctgctctagaatgagccacacagaagatcgtttcaagagcccaagatcgtggat 381  
|||||  
Db 181 TCTGGCTCTAGAAATGACACCAACAGAAATGATTCAGGAGCCGAGAACTCTGGGAT 240  
QY 382 atgtctctgagatgagtgacagaacacacatcactggtctgtgtgactcaatgac 441  
|||||  
Db 241 ATGTTCTTGATGAGCTGTGACAGAAACATCACTACTGTTCTGTGATCCTCATGAC 300  
QY 442 atacttggtgagcggatcactgcaagaagatcactgagatcgtgtggtcccaagggt 501  
|||||  
Db 301 ATACTTGTGGCGGATGATCAGTCAAGAAAGATTACTGATATCGGTGGCGTCCAGGGGTT 360  
QY 502 ggttaaaatcactggtggtatcaactagcaatcaatgtcaaatcccaatggaatgtgt 561  
|||||  
Db 361 GGTAAACATCACTGGGGATTCATACAGCAATCAATGTAATCCCATGTGAAGTGTG 420  
QY 562 ggccttggtggaagacgttatataatagatacagaagggagttcaatgtgtgaagctgc 621  
|||||  
Db 421 GGCTTGGTGGGGAACACATTTATATAGTACAGAGGCGAGTTTCAATGTTAAAGTGTG 480  
QY 622 taccagatctgtaaggtgtgattaggaacactggaagcattccgcagacagatgag 681  
|||||  
Db 481 TACCAATGTTGTAAGAGGGGTATAGGAGCATATGAGACACTTCCGACAGCATGAG 540  
QY 682 aagtcctctctgtccaaaacaaatatacagcgtgagcgtttcctgtgagatactatc 741  
|||||  
Db 541 AAGTCTTCTGTGTCACAAACAAATTCAGCTGAGCGCTTCTGCGGATATCTATTAC 600  
QY 742 ttccgatatgcaatcgcgaacaaatgcaatgcaatcaatcaatcaatggaatgtctc 801  
|||||  
Db 601 TTCCGATATGCAAGTTACACGCAACAAATTCAGATCATTAATGAGAGATTTCTC 660  
QY 802 agagagcataaagatgtgcgtatagttatattatgattagtgattcttccacttcgaca 861  
|||||  
Db 661 AGAGACATTAAGATGTGGGTATAGTTATTGATAGTGTACTTCCACTTTCGACAA 720  
QY 862 gatttgaagatctgagcactgagagacagagtgctcaagtgtgattatcaattgaagt 921  
|||||  
Db 721 GATTGTAAGATCTGCGACTGAGAGACAGATGCTAAGAGATTTATCATGTAAGTTAATG 780  
QY 922 aagattgcaagaacataataacttggcaatgtgtctgttgaaccaagtcaactaatt 981  
|||||  
Db 781 AAGATTGCAAAACATATATACCTTGGCAGTTGCTTGTGTAACCAAGTCACTATAATT 840  
QY 982 acagaaggtcattcaattgactctgtctaggttgacagctgtgtccactactgacag 1041  
|||||  
Db 841 ACAGAAGGGTCAATTTCAATTAAGTCTGTGCTTAGGAGACAGTGTCCCACTCATGACG 900  
QY 1042 aacgggttgaatctgactggaatggaagacagatacgcacatcttgataagctctc 1101  
|||||  
Db 901 AACCGGTGATTTCTGCACATGGAATGGGAACGATACGCACTGTGATAGTCTCTCT 960  
QY 1102 tcaattcagtagctcagcaccgctatcgaatggaagcaaggaataagagaatctgtg 1161  
|||||  
Db 961 TCACTTCCAGTAGCTTCAACACCGATGATGACAGGAAAGGATTAAGATGCTGTG 1020  
QY 1162 agtcaaacacagaagccgagtaacgtacattctgtgtgcaagcactgtatgt 1221

Db	1021	AGCTAAACGACAAACGAGCCGAGTAACGTATGACTTCTGGTGTCAAGCACTTGTATGT	1080
Qy	1222	ccactagctcctcgcagctcttccttcgcataggatccttttgactcagtgagtgagcttgg	1281
Db	1081	CCATCAACCTCTCTCGTGTCTTCTTCGTGCATGATGATTTTGGCATGTAGTGAAGGAAACCTGG	1140
Qy	1282	agaatagcaccattcttgctgattctcagttgtctgtgcccgttgcctaccacccaacct	1341
Db	1141	AGAAATGATGCCATTTTGTGTGATTCCTCACTGTGTTGTGCGCTTGGCTACCAACCAACTTT	1200
Qy	1342	aagagagaagtaataacacacagaaacaggtcataatagtglttltglatctgaaacatctggc	1401
Db	1201	AAGAGAGAAAGTAAATATCAACAACAGACAGCTAATATAGTGTTTGTATCTGAGCAATCTGGG	1260
Qy	1402	ccatcgtatcattcagttaaagctataatctagcggcgaataaaaaaataaaaaaataaaaaa	1461
Db	1261	CCATGTCATCTTCATCAATAAGCCTATTAATATGCGGGCATATATGTGCTTCTGTGATCAAAA	1320
Qy	1462	aaaaaaaaaaaaa 1474	
Db	1321	AAAAAAAAAAAAA 1333	

Db	121	GTTCCTCTTCTCCCAATAGTACCACTCTCCAAAGTGCAATCACCATGGAGATCAA	180
Qy	322	tcctggcctctagaatggtaccacacacgaagtgcgttctaaaggagcccaaatgcttggtat	381
Db	181	TCCTGGCTCTCTGAAATGAGACCAACAAAGAACTGCTTCAGGAGCCCAAAATAGCCTGGGAT	240
Qy	382	atgtctctctatagcctgctcacagaacaacacatcaactactggtcttcgtgacccatgac	441
Db	241	ATGTCTCTGATGAGTGTCACAGAAACACATCACTACGCTGGTTCGGTGGACTCCATATGAC	300
Qy	442	atactctgtgtgtgggattcaactctgcacaagaagttaactgtagatcgtgtgctgccagggttc	501
Db	301	ATACTTGGTGGCGGGAGTTCTACTGTCCAAAGAAAGTTACTGAAATGGGCGGTCCACGGGGTT	360
Qy	502	ggttaaacctaactctgtgggtatcaactacagaaatcaatgtgcaaatccagtggaatgtgt	561
Db	361	GGTAAACTCAACTGGGGATTCAACTTAGCAATCAATGTTCANATCCCACTGGAATGTGTGT	420
Qy	562	ggcctctgtgtgaagacagtttatatagaatacagaaggcgagttcaatggttcgaacgtgtc	621
Db	421	GGCCTTGTTGGGAAAGCAGTTTATAT-----AGAGGGCAGTTTCATGCTTGAACGTGTC	474

LOCUS	AX046603	1459 bp	DNA	PAT	15-DEC-2000
DEFINITION	Sequence 3 from Patent WO0068370.				
ACCESSION	AX046603				
VERSION	AX046603.1	GI:11876175			
KEYWORDS					
SOURCE	Zea mays.				
ORGANISM	Zea mays.				
	Eurycotla, Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea.				
REFERENCE	1 (bases 1 to 1459)				
AUTHORS	Mahajan, P. B. and Shi, J.				
TITLE	Rad51 from maize				
JOURNAL	Patent: WO 0068370-A 3 16-NOV-2000;				

[illegible]

FEATURES	source	location/Qualifiers
		1..1459
		/organism="Zea mays"
		/db_xref="taxon:4577"
CDS		169..1014
		/note="unnamed protein product"
		/codon_start=1
		/protein_id="CAC18902.1"
		/db_xref="gi:11876176"
		/translation="MGDSSGSRNGPOOKYYSGAONAMFSEDELSOKHITTGSGLIN
		ILGGHCKKEVTEIGVPGVGVKTQIGIDALNVQIPVCGGELGCAVIEESFVER
		YQIEGCIKDLEHPPHSEKSSSVQKQLOPERLADIVYRISGYEQIAVINMER
		FLRHKDVRIVIDSVTFEHRQDEFLDLRTRVISGLIKLMAKTAKTVNLAVVLNG
		TTKTEGSGFQLTLALGDSMSSCNTRLILHMNGERYAHLDKPSLPAVASAPYAVTGS
		GIRRV"
BASE COUNT	504 a	288 c 326 g 341 t
ORIGIN		

Db	715	GATTTCGAAATGTCGGCAGCTGAGGACCAAGAGTCGTAAAGGATTTTCATTGAAAGTTAATG	774
Oy	922	aaagtttcaagaagcatalaactcttggcagtttgccttttgaaccaagtaactacacaaattt	981
Db	775	AAGATTCGAAGACATPATACTTGGCAGATTGCTTGTATACCAAGATCACTACATPAAATTT	834
Oy	982	acaaagaagtcatttcaattgaactcttgccttagtgaagccttggctccactcatalgcag	1041
Db	835	ACGAAGAGGTCATTTCATTTGACTTTGGCTTAGTGAGACAGCTGGTCCCACTCATGACAG	894
Oy	1042	aaacggttgaattctgcactcgtgaatbvggaagaaagataagacacatcttgataaagtctct	1101
Db	895	AACGGTGTGATTTCGCACCTGGAATGGGAACGAACGATPAGCAATCTTGATPAAAGTCCTT	954
Oy	1102	tcacttcacgaagcctcaagcaccgtatgcaqtgacagcgaagggattagaagatctgctg	1161
Db	955	TCACTTCCAGACGCTCAGCCCCGCTATGCAGTGCACGCAAGGAGATTAGAGATG-TCGT	1013

	Query Match	Similarity	84.3%	Score 1243.2	DB 9	Length 1459
	Best Local	Similarity	98.9%	Pred. No. 8.1e-290		
	Matches 1286	Conservative	0	Mismatches 3	Indels 11	Gaps 3
Qy	142	cgaagtaacgagctgctgtgcgcacacgagaggtctacagacggtttgtgagcgagata	201			
Db	1	CGACGTAAcGGGCTGCGTGGCCCCACGACGAGAGGCTACGACGGTGTGGAGGCGAGATA	60			
Qy	202	tgaagagtgtagaggttggtctacaacgaggttcgagcggtctgtgatactgtaaatccgcacttga	261			
Db	61	TGAGAGGTGTAGAGGTGGCTACAAcGGGTCGGCGGCTGTGAGATACGTGAATTCGCACTGCA	120			
Qy	262	gttctctcttcttcccccacatagtaaccacctctccaagtgtgcaatcaatcattgtagagatcaa	321			

Db	1014	AGCTCAACCAACGAAGCCGACAGTAAAGTAGACATTCTTGGTGTCAAGCACTGTATGT	1073
Qy	1222	ccactacagctctctgcagcgttctcttcgcacatgatctctttggaactagtgagtgtagactcg	1261
Db	1074	CCACTACGCTCTCTGACACTTCTTCCGCACATGATCTTTGGACTAGTAGAGGAGACTGG	1133
Qy	1262	agaatagatcacatttgcttgatcttcaagttgcttcttgccgttggtctacaacaaactt	1341
Db	1134	AGAAATAGTACCAT---TTATATCTCAGTGTGCTTGTGGCTACCAACCAACTT	1189
Qy	1342	aagagagaagtaaatatacaagaacagcgctaaatatgcttctgtatctggaacatctgac	1401

```

Db 1190 AAGAGCAAGTAATACACAGAGGCTAATATAGTCTTTGTATCTGACATCTGCC 1249
Qy 1402 ccatcgtacattcagtaaacgctataatagcgagcaaaa 1441
      |||||||
Db 1250 CCATCGTACATTCAGTAAAGCCTAATADAGCGGCATATA 1289

RESULT 4
ATH299424 1101 bp mRNA PLN 20-OCT-2000
LOCUS Arabidopsis thaliana mRNA for RAD51C protein.
ACCESSION AJ299424
VERSION AJ299424.1 GI:10944744
KEYWORDS RAD51C gene.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
          Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
          Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
          Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Slaud,N.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1101)
AUTHORS Slaud,N.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-2000) Slaud N., Reparatoin/Recombinaison de l'ADN
          630, Universite Paris XI 91405 Orsay Cedex, FRANCE
          Location/Qualifiers
            source 1..1101
                  /organism="Arabidopsis thaliana"
                  /cultivar="Columbia"
                  /db_xref="taxon:3702"
                  /chromosome="II"
                  /gene="RAD51C"
                  /gene="RAD51C"
                  /protein_id="CAC14091.1"
                  /db_xref="GI:10944745"
                  /translation="MISFGRKRKSPAIETSLATSVYMWAMPLPSIRKLIAGTYC
                  /LSTASVSSDLDANITEFAETILKLNQSCNSRLINGAKNAMDLPBESL
                  /PRITSCSDLDNIGGIGISCRDVEIGVPGIGTQIGLQSVNVDIPREGSLGGA
                  /IYIDTESFWMKALQIAEACVEDMEYEMRHFQANOVOMKPEDILNIFYPRVC
                  /SYEOLALVNHLEKFISENKDVKYIVDSITFHRPDYDLAORTRVLSMALKFMKL
                  /AKKPSLAVLLNVTTKFESGSPOLATALGDSMSHSCITNRVILYWNDEYAYIDKSP
                  /SLPSASAYTTSRGLRNSSSSKRYKMM"
BASE COUNT 307 a 219 c 268 g 307 t
ORIGIN

Query Match 24.3%; Score 358.8; DB 13; Length 1101;
Best Local Similarity 65.7%; Pred. No. 3.8e-76;
Matches 522; Conservative 0; Mismatches 272; Indels 0; Gaps 0;

Qy 361 ggaagccagatgctggatgattgtctctgtagtgagctgtcaagaacacatcactact 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 262 GGAGCAAGAACGCTGGATGCTTCACAGAGAGAGTCTTCCCGATATCTACACA 321

Qy 421 ggtctgtgagactcaatgacatctactgtgctggagattcactgtaagaagattactag 480
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 322 TCTTGTCTGATCTTGATACATTTTGGCGGTGGAATTAAGCTGTAGGAGATTACAGAG 381

Qy 481 atcgtgtgcctccagggggttgtaaaactcaactcgtggatccaatcagcaatcagtla 540
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 382 ATTGGTGGGTACACAGGAGTGTGCAAGACTCAGATTGGATCCACCTCTCTGTGAATGTT 441

Qy 541 caatccagtggaatgtgtgcttggtaggaagcagattatagatagaagaagc 600
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 442 CAGATTCACGTCGAGTGTGTGTGAGGGAAGATATATATATCATACAGAGCT 501
Qy 601 agttcatgttgaacgtgtctaccagattcgtgaagggtatattaggaactactgag 660
      ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 502 AGCTTCATGTGTGAAGGCTCTTACAGATAGCAGAGAGCTTGTGTAGAGCATGAGAA 561

Qy 661 cacttcgcagagcagatgagagtcctctctctgtccaaacaaatlaagcctgagct 720
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 562 TACACAGATACATGTCATTAACATTTTCAAGCAAAATCAAGTACAAATGAAACCAAGAT 621

Qy 721 ttcttgcgatalctcattactcctcgatatacgtatcaccgaaacaaatcagtcata 780
      ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 622 ATCTTGAGAACATATTTACTTCCGTGTCAGTTACACCGAGCAAAATCCGATTTGGTC 681

Qy 781 aactcatgagagaagttctctcagagagcataaagatgtgtcgtatagttattatagt 840
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 682 AATCATCTTGAAAGATTCATCTCTGAAACCAAGATGTTAAAGTTATTCGTAGCAGCT 741

Qy 841 gttacttccacttcgacaaagatttgaagatcgtgcaactgagagcagagtgctaagt 900
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 742 ATCACCTTTTCATTTCCGTGACGACTATGATGATTAAGCCAGAGGACAGAGTCTCAGC 801

Qy 901 ggaattcatgtaagtaatgaatgcaagatgcaagacataactgtgcaattgtctgtg 960
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 802 GAATGGCTTTTAAAGTTTCATGAAGCTTGCCAAAAGTTCTCACTTCGCTGCTGTTACTA 861

Qy 961 aaccaagtcactcctaatttaccagaaggtcatttcaatgactcttctcagtgagc 1020
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 862 AACCGAGTGACCAAAAGTTTACTGTAAGGCTCTTCAACTAGCCGTTGCTTTAGCGCAT 921

Qy 1021 agctgtcccaactcatcagcaaacggtttagtcttctcactggaattggaagaacatgc 1080
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 922 AGCTGCTTCATTCGTGACCAACCGAGTCACTCTGATTTGAATGATGATGATGAGCTTAC 981

Qy 1081 gcaatcttgataagtcctcttacttccagtgagtcctcagcagcagatgagtgagc 1140
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 982 GCATATATCATGATTAAGTCCCTTCACCTTCTTCACTTCGCTGCTTCACTACCTGATACCAGT 1041

Qy 1141 aaaggagattagaga 1154
      ||||| |||||
Db 1042 AGAGGCTTAAGAAA 1055

RESULT 5
AF029669 1295 bp mRNA PRI 24-FEB-1998
LOCUS Homo sapiens Rad51C (RAD51C) mRNA, complete cds.
DEFINITION AF029669
ACCESSION AF029669.1 GI:2909800
VERSION
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1295)
AUTHORS Dosanjh,M.K., Collins,D.W., Fan,W., Lennon,G.G., Albaladejo,J.S.,
          Shen,Z. and Schild,D.
TITLE Isolation and characterization of RAD51C, a new human member of the
          RAD51 family of related genes
JOURNAL Nucleic Acids Res. 26 (5), 1179-1184 (1998)
MEDLINE 96136197
REFERENCE 2 (bases 1 to 1295)
AUTHORS Schild,D., Collins,D.W. and Dosanjh,M.K.
TITLE Direct Submission
JOURNAL Submitted (09-OCT-1997) Life Sciences, Lawrence Berkeley National
          Laboratory, Ms. 70A-1118, Berkeley, CA 94720, USA
FEATURES
  source 1..1295
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="17"
        /map="17q:413.6cR from the top"
        /cell_type="Leukocyte (mixed population)"

```



Db 2399 TTTCCTGACAAATTAGTCTGATATTATGTTT 2433

RESULT 7  
AP002908/c 163419 bp DNA PLN 27-FEB-2001

LOCUS  
DEFINITION  
Oryza sativa genomic DNA, chromosome 1, PAC clone:p0013602.  
AP002908  
AP002908.2 GI:13161334

KEYWORDS  
SOURCE  
ORGANISM  
Oryza sativa (cultivar:Nipponbare) DNA, clone:p0013602.  
Oryza sativa  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;  
Oryza.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Sasaki,T., Matsumoto,T. and Yamamoto,K.  
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
clone:p0013602  
Published only in Database (2000) In press  
2 (bases 1 to 163419)  
Sasaki,T., Matsumoto,T. and Yamamoto,K.  
Direct Submission  
Submitted (08-NOV-2000) Takuji Sasaki, National Institute of  
Agrobiological Resources, Rice Genome Research Program, Kannondai  
2-1-7, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tsasaki@abr.affrc.go.jp, URL:http://ryp.dna.affrc.go.jp/  
Tel:81-298-38-7441, Fax:81-298-38-7468)  
On Feb 28, 2001 this sequence version replaced gi:11136550.  
Genes were predicted from the integrated results of the following:  
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor  
(October 1998 version). The genomic sequence was searched against  
NCBI NonRedundant Protein database, or  
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at  
RCP. Protein homologies of the coding regions were searched against  
NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent  
the identified cDNA sequences using BLASTN 2.0 with the  
corresponding DBS accession no. and RCP clone ID.  
A gene with identity or significant homology to a protein is  
classified based on the protein name to indicate the homology level  
such as same name, 'putative' and '-like protein'. A gene without  
significant homology to any protein but with EST homology (covering  
almost the entire length of partial sequence) is classified as an  
'unknown' protein. A gene predicted with a gene prediction program  
is classified as a 'hypothetical' protein.  
The orientation of the sequence is from 5' to 3' of the PAC clone.  
Detailed information on overlap and assembly quality together with  
annotation of this entry is available at  
http://ryp.dna.affrc.go.jp/genomeseq.html.

FEATURES  
source  
1.163419  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/chromosome="1"  
/clone="P0013602"  
complement(join(2332..2436,2786..2871,3194..3197))  
/gene="P0013602.1"  
complement(join(2332..2436,2786..2871,3194..3197))  
/gene="P0013602.1"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAB32926.1"  
/db\_xref="GI:13161335"  
/translation="MDLAKGYTSPPTRCGEGEDSSLWLEKMRMRKRGILLPVGVD  
EERKPKRTASVLEGVNRH"  
complement(8102..9679)  
/note="3' LTR"  
/gene="P0013602.2"  
complement(9637..12886)  
/gene="P0013602.2"  
/note="probably inactive due to stop codon(s) in CDS

gene  
CDS  
pseudogene  
similar to polypeptide  
complement(join(13135..13944,14966..15037))  
/gene="P0013602.3"  
complement(join(13135..13944,14966..15037))  
/gene="P0013602.3"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAB32927.1"  
/db\_xref="GI:13161336"  
/translation="MSWMSAGVLLGAOSCLPILVPMAGSVNRHGLDLCFVLELR  
GMCIVLGPHEVDYOGKKLPEDODDAEPHAMVAVIILTSPERTSLAVTAGGSF  
SAACONALLAIGTLHOHYPELDGHSRYRHPRRGADATFMDASDDATMHLA  
CMVAYDARIDFHHMYRRCIVENNMILTELOKDKDVADEAOLFOIKAOGE  
VCHRRRRVCSLKITARKSTFPRELVRSLAMTCFVETPCAPAPVPODEBSGSG  
STEVLLTFEPSPQR"  
complement(15279..16871)  
/note="5' LTR"  
join(16203..16463,18333..18929,19065..19457,20680..20751)  
/gene="P0013602.4"  
join(16203..16463,18333..18929,19065..19457,20680..20751)  
/gene="P0013602.4"  
/codon\_start=1  
/product="putative transcription activator"  
/protein\_id="BAB32928.1"  
/db\_xref="GI:13161337"  
/translation="WVGGGHRRHGAAGAECEKEETRRNSTAMPSTATTKRAAGV  
EGGGAARDLDGAPALGEDEVDGDCARKEEATPGEAASEPNNVNRITAF  
WEILPACSDPDISGTPSPSPPTGEGDANVTVIDFLAHSIDLPAAGDQKKE  
HOLDLDGQQQAAMEIDDVYIISLRNMNDSSSSIKINMERRRRKLEMFIL  
TELKKEAVAGEAORLLKRVGGGAGMAASCSIKINMERRRRKLEMFIL  
LKSIVPSIDKKAISLSETIAYLKELERRVQDELESKRPPKRCSETIIGGG  
GGGAGAVKEHHNVLSSESGTSDVRYVQDELEHVOCKKRLMTFRFPAIK  
SLRDVVSVOASAPDGLGKTRKRYASNAVPAVAMSETLRTVAGY"  
join(28520..28527,29674..29780,30119..30294)  
/gene="P0013602.5"  
join(28520..28527,29674..29780,30119..30294)  
/gene="P0013602.5"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAB32929.1"  
/db\_xref="GI:13161338"  
/translation="MTLEPRGRGMRRCSGMETMPAEVPECGATREGGRTARQ  
GIDALDCGDSNGHDDSGRLLLAHLAAGLPLPSTSPAAPCYFVCV"  
complement(join(31561..32021,32574..33461,33560..33820,  
34399..34658))  
/gene="P0013602.6"  
complement(join(31561..32021,32574..33461,33560..33820,  
34399..34658))  
/gene="P0013602.6"  
/codon\_start=1  
/product="putative receptor-like protein kinase"  
/protein\_id="BAB32930.1"  
/db\_xref="GI:13161339"  
/translation="MAASRLVAAAIIVVVLVLAGAAGAAGAEVALLDLAAGLD  
PTGRLIPSMAGRDPCGEGGEGGEGVAGADATAVANVSLQGLAGTLPPVAGLT  
ALTGLYLHNRITGALPRELAALSRLDVLKANNVSGPIPEIAPALPAGSLP  
OLTGVPVQDLKRLTYLEQSNHLSAIPASGDI.PQVLYRLDLSFNNLFGSIPVRL  
ALLPRLALDVRRNNTLTGSVSELAIDGGQVQANNDVLAACVAVIILTAAGMFA  
DMPOFSGISPOITPGSSSDGHCCTGHPSTKALAAVAVVAILTAAGMFA  
LSWFKRKORVAVGSPPAVAGRCSTDAAGDLPFRKASSTLVLSLEYSNGMDPLADGGG  
GIGFEOEAVOSFRENMEDVESATQYFESLNLTKNGNFAATYRGLTDGTSVYVRRLG  
KTCKKOEAEFLKGLIALEOHENIYGLIRFSCSRAGCEFLVYDVPNGSLSDPFD  
IDADVAASRNVLEWRSRIITIGIKAGIYGLHSTANRKEPLVYDHONISADKYLVDY  
YRPLISGGLKRLIYDLVFTSLKASAMGLAEYTTTGFSSEKSVYARCVYVPOI  
LTGSKIMQPLESSNDEDLIDGNLRGCTSAEAKLAKIASCTSENPHRPMEL  
IQDELTFP"  
complement(join(47103..47281,49010..49076,49786..49917,  
50114..50340,50420..50490,50561..50594,50835..50957,  
51241..51345,51594..51705))  
/gene="P0013602.7"  
complement(join(47103..47281,49010..49076,49786..49917,

5014. .50340,50420. .50490,50561. .50594,50835. .50957,  
51241. .51345,51594. .51705))  
/gene="P0013602.7"  
/codon\_start=1  
/product="putative Rad51C protein"  
/protein\_id="BAB32931.1"  
/translation="MEIADLPATSHRANLAIAGYSIALSAASAPRLARDLSEVH  
EAEILKAVAGANKSGADGPTSVLGVONAMDMDESRHINIGSDLNING  
GGIHCKEYVEIGVPGVKGKLTGLOLATNOIPYEGGIGGKAYIDGESFEMERY  
QIAECSIDILEYPPHCHDKAPAGEKIKPSPLADIIYFICSTBEDINANTLEK  
LGEHKDVRIVITDSVTFHFRODFDMALRTVLSGLSLKLKLSKANVAVLLNOVT  
TKTEGSEFQLTLALGDSHSCNTNRLIYMNNGERYGFLDKSPSLPVASAPYAVTVG  
VDAVANSKSRKRVW"  
join(53315. .53371,53419. .53458,53464. .53574,53856. .53948,  
54151. .54511,54550. .54925,55256. .55423,57768. .57953)  
/gene="P0013602.8"  
join(53315. .53371,53419. .53458,53464. .53574,53856. .53948,  
54151. .54511,54550. .54925,55256. .55423,57768. .57953)  
/gene="P0013602.8"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAB32932.1"  
/db\_xref="GI:13161341"  
/translation="MDPRAFLGRPIRMNRASGDEGEMCLIKAFALAFGRSSSLVNP  
LTFVARLRPRDKPLSLPSGFAQVLSHGVRVYSAIYCAKSDRNLFRLPLVTLAS  
LSPLRISRLANLVLGOAKTIGASPTTELATWASHSKMPGMLAGVTEGGA  
SNPVDDEVORSGDKRTVYFAAFDAGRLCLDDFLASVLEIYEARLPOLSLATKSD  
VETARFAFGCOREDEGRTHACRTKEIGVRQROPPGARGFLAGSGIHKTSKIDS  
AEMGLYINPTAKYSDNAGRLRPRCLPISINSKPSMITITLESRIILANVCRLS  
THDCCEYCMLORTORAGEHPCOLNHLHFSAGDRLATLDMTESCLMCLDQLER  
MYLFKCLDLAGEISTESRTICQNFVIGILFLANRNGFVRIADKKTOLAVSIQTAR  
OCNRSSTNRTION"  
complement(join(58887. .59081,60870. .61024,61113. .61203,  
61319. .61416,62180. .62297,62345. .62470,62559. .62774))  
/gene="P0013602.9"  
complement(join(38887. .59081,60870. .61024,61113. .61203,  
61319. .61416,62180. .62297,62345. .62470,62559. .62774))  
/gene="P0013602.9"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAB32933.1"  
/db\_xref="GI:13161342"  
/translation="MEPTDKLQSLQINPNKNTNLVAVFNKNDYVDEDDMDPOVTL  
GTEIEPEPEDEMHLLPQHPFNKAGGVPAWLPVNLPSGKRCDCDFGEPEFLVLYV  
FHLXDLQVNOVVALIOPKETAHRTLFVPCSMACILIDHOEGEDRAGSTWFC  
CMWGKGEKYSCKRSKSCSKKHQHEHMRKHKNEHQHSGSHNSAIMPDKGXF  
AGNITPEYAVYDETEKYSKSCFASCEKRSFLMEGOCSEDDTASLMQFETMDPLLH  
FHVNEPDSLMAITIVITCKGSCDQNSIMEEFVWOLSPATRTNOSTCPSAGL"  
join(63427. .63486,63492. .63811,63861. .64053)  
/gene="P0013602.10"  
join(63427. .63486,63492. .63811,63861. .64053)  
/gene="P0013602.10"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAB32934.1"  
/db\_xref="GI:13161343"  
/translation="MALTSAMTSSPATSKGAPRROHREHREGGDGLTNGGDE  
KRRETAGARCPWMNDPATSQNGVAGRLDAAMPKKEETVITSDDNGGATAHSG

Query Match 11.6%; Score 171; DB 13; Length 163419;  
Best Local Similarity 76.4%; Pred. No. 1.2e-30;  
Matches 210; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

OY 581 ttatataagacagagcaggttcacatggttgaacgtgtctaccagattgtaaggt 640  
DB 50349 TTTTAAACAGATTACAGAGGAAAGTTTCATGTTGAACGTTCTACCAAAATGCTGAAGGT 50290  
OY 641 gtattaggaactactgagacacttccgcacagcagatgagacccctctctcccaa 700  
DB 50289 GCATCAGTATTAAGTGGAGTACTTCCACACTGCATGACAGAGCTCCAGCGGACAAAG 50230  
OY 701 aacaattacagcctgagcgcttctcgtcgagatattacttccgagatgacattaca 760

DB 50229 AAAAAGTGAAGCTGAGAGGTTTCTGGCTGACATCTACTACTTTCGAATATGACAGCTATA 50170  
OY 761 ccgacaacattgcagctataactaactatgagaaggtccaccagagacataaagatgtgc 820  
DB 50169 CTGAACAAATTCAGTCAATAAATATCTGAGAAAGTTTCTGGGAGCATTAAGATGTAA 50110  
OY 821 gtacattatcatcatgatagttctacttccacatt 855  
DB 50109 TTTTCTGAACAAATTTAGTCTGTAATTTATTAAGTTT 50075

RESULT 8  
BC000667  
LOCUS  
DEFINITION  
MGC:2023, mRNA, complete cds.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT

BC000667 1123 bp mRNA PRI 16-MAR-2001  
Homo sapiens, similar to RAD51 (S. cerevisiae) homolog C, clone  
MGC:2023, mRNA, complete cds.  
BC000667 GI:12653762  
MGC.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1123)  
Strausberg, R.  
Direct Submission  
Submitted (15-NOV-2000) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nih.gov](mailto:nisc_mgc@nih.gov)  
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Barkley, C., Brooks, S.,  
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
Lim, M., Maduro, Q.L., Masello, C., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Snyder, B., Stantirpop, S., Thomas, P.J.,  
Tongson, E.E., Touchman, J.W., Tsurgueon, C., Vogt, J.L., Walker, M.A.,  
Zhang, L.-H. and Green, E.D.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAL Plate: 5 Row: d Column: 24.  
Location/Qualifiers  
1. 1123  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:2023 IMAGE:3350026"  
/tissue\_type="Eye, retinoblastoma"  
/clone\_lib="NIH\_MGC\_16"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
39..443  
/codon\_start=1  
/product="Similar to RAD51 (S. cerevisiae) homolog C"  
/protein\_id="AAH00667.1"  
/db\_xref="GI:12653763"  
/translation="MRGKTFREMDLVSPFSPAVRYKLVASGFGTAELLLEKPS  
ELSKVEIGSKALELTLITIRRECLTNKRPVTSBSHKKTALLLDQEHQGFIIIT  
FCSALDILIGGVPLMKTIETICGAPGVGTQL"

BASE COUNT 346 a 226 c 255 g 296 t  
ORIGIN



SOURCE  
ORGANISM  
thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 122871)  
Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D., Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnes, M.E., Feldy, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M., Koo, H., Moffat, K.S., Cronin, L.A., Shen, M., Vanaken, S.E., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Coppenhaver, G.P., Preuss, D., Nieman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C.  
Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana  
Nature 402 (6763), 761-768 (1999)

JOURNAL  
MEDLINE  
20083487  
PUBMED  
10617197  
REFERENCE  
2 (bases 1 to 122871)  
AUTHORS  
Lin, X.  
TITLE  
Direct Submission  
Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA  
On Dec 17, 1999 this sequence version replaced gi:2583106.  
The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (<http://www.tigr.org/tdb/at/at.html>).

COMMENT  
Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (<http://arthur.epm.ornl.gov/pub/grail/>), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://genomic.stanford.edu/GENSCAN.html>), and NecPlantGene (<http://www.cds.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the GSHL/Mashu/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T1316, the ESSA group for sequencing clone F13B4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalek, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: [atetigr.org](mailto:atetigr.org).  
Location/Qualifiers  
1. 122871  
/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="II"  
misc\_feature  
complement(<1..2952)  
/note="Sequence from clone T14P1"  
repeat\_region  
complement(519..627)

mrna  
/rpt\_family="(TAAA)n"  
join(2227..2322,2480..2527,2650..2705,3098..3150,3250..3641)  
/gene="At2g45170"  
2227..3641  
/gene="At2g45170"  
/note="T14P1.2"  
join(2488..2527,2650..2705,3098..3150,3250..3469)  
/gene="At2g45170"  
/codon\_start=1  
/product="putative microtubule-associated protein"  
/protein\_id="AA18602.1"  
/db\_xref="GI:6598366"  
/translation="MNKGSIFKMDDEPKRKAERIRKRPDRIPVYKAEKSEVP NIDKKYLVPSDLTGQFVYVIRKRIKLSAEKALFIFVNVLPPTGELMSVYEDKKD EDGLVITYSGENTFGASST"  
complement(2953..122871)  
/note="Sequence from clone F4L23"  
4282..4870  
/gene="At2g45180"  
4282..4870  
/gene="At2g45180"  
/note="F4L23.31"  
4320..4724  
/gene="At2g45180"  
/codon\_start=1  
/product="putative proline-rich protein"  
/protein\_id="AA82643.1"  
/db\_xref="GI:2583134"  
/translation="MSKALAVYALITLNLFFETVSTKCPPTTPKPKPKPSKK APVAKPCPTDITLGLVADLGLVNVVSGSPKPCITLGLANLEAAVCLTALK ANVLGINLVPIIDITLLNLVCGKRPVHGFQCS"  
5379..5403  
/rpt\_family="POLY A"  
complement(join(<6393..6470,6586..6663,7082..7157,7248..7296,7872..8004,8099..8260,8382..8495))  
/gene="At2g45190"  
complement(<6393..8495)  
/gene="At2g45190"  
/note="F4L23.30"  
complement(join(6393..6470,6586..6663,7082..7157,7248..7296,7872..8004,8099..8260,8382..8495))  
/gene="At2g45190"  
/note="unknown protein"  
/codon\_start=1  
/protein\_id="AA82644.1"  
/db\_xref="GI:2583135"  
/translation="MSKSSMSPSAASVSPDHPSPDHLCTVOCNFCOTILAVNPEYTLRTVYVRCGCTNLISVNRBYVLPASMOLOIGPSHYEPDILIELDAPSNM NMAMNQHPTMNDIPSPMDLHOCHETPKAPVVRPEPKQRPSTAIRITKEIQRIK AGNPDISHREAFSAAKNMAHPHIFHGLVPDQPVKTNMPQEGEDMMVMEGFTY PAAANVGVPY"  
10147..10201  
/rpt\_family="(GA)n"  
12613..12646  
/rpt\_family="(CAT)n"  
complement(join(15613..15729,15821..15950,16066..16137,16252..16341,16810..17006,17451..17647))  
/gene="At2g45200"  
complement(15613..17647)  
/gene="At2g45200"  
/note="F4L23.29"  
complement(join(15632..15729,15821..15950,16066..16137,16252..16341,16810..17006,17451..17583))  
/gene="At2g45200"  
/codon\_start=1  
/product="putative cis-Golgi SNARE protein"  
/protein\_id="AA82642.1"  
/db\_xref="GI:2583133"  
/translation="MTESSLDLQESGMEELRREARKIEGDLVYKSSVYKLGARFTQG GYVDGSPVGVSGSWKSMEMETOSLEKLDINDSMRCAASAAPTSTVQLAKRH DILHEYTOEFFRIKGINSLREAEALISVARDISYKASGSMSPGVQLRRASIHG SISHIDVIOGAQATRAVLGSGRSLFSDVQGVKYNKGDFPVIRGLLSIKRRSMDT

```

misc_feature
complement(17986..18862)
/Note="molecular marker ML"
repeat_region
/rpt_family="(TAAA)n"
19741..>20315
/gene="At2g45210"
19741..>20315
/gene="At2g45210"
/Note="F4L23.28; member of auxin-induced protein multigene family"
19827..20315
/gene="At2g45210"
/codon_start=1
/product="putative auxin-induced protein"
/db_xref="GI:2583132"
/translation="MKRIIGFRIGRRYRNIFKRTIORSYNRHISTOQACMLMPLAKLSWGQRLKQSFRRSRSAIIPVDHKAQDPVPGHAIIVGQKDGCHVLVPIVFNHPLFGLLEAREKEKGFCHEGGITIPCLSDPERVKTIRIASGSSSRVFPWGRHC RN"
mRNA
complement(join(<22224..22909,23488..>24334))
/gene="At2g45220"
complement(<22224..>24334)
/Note="F4L23.27"
complement(join(22224..22909,23488..24334))
/gene="At2g45220"
/codon_start=1
/product="putative pectinesterase"
/protein_id="AAB82640.1"
/db_xref="GI:2583131"
/translation="MAFRAYIINFEVILICILVASTVSGYNOKDYKAWSCPNPCEY/FLTNSNNEPIKSESEFLISMKLVDRALATKATFTLGPCKRDRBKAAMDCKLVDLTVSKINFTMDPNVYCKSLDAOTMLSTALDLCRAGLELGTVDIYPLMSNVSNLCNTLAINKVPNTTPPEKDPSPWAKGGRKLIQSTPDKNVAADSGNKTKKEAIDASGSGRVIYKQGVYSENLEIRKNVMRGDGIKTIITGSKSVGGTTFNSATVAAGDGFJARGITERNATGASNEQAVLRSGSDISVYQCSFEAYODTLVHSNRFYRDCDVGTFDEIFGNAAVLQNCNIFARRPSKNTITTAAGRSPPDNLTIIINHSRVTAAASDLRPLVSGSTKTYLGRFWRVSRTVPFKTISLSDIDRGMLEMGNALKTLEFAFQNTGPGASISGRVTFGRFVLSASBAKFTYGTFLAGSGNIPSSVPTTS GL"
repeat_region
23009..23038
/rpt_family="(TA)n"
repeat_region
28701..28741
/rpt_family="POLY_A"
complement(29132..29228)
/rpt_family="POLY_A"

```

Query Match 6.9%; Score 101.4; DB 12; Length 122871;  
 Best Local Similarity 59.0%; Pred. No. 7.5e-14;  
 Matches 174; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

```

585 tatagatacagaagcattcatggtgacggtgtaccagattgctgaaggtgat 644
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
49087 TACAGATTACAAAGACTTCACTGAGGACCGTCTTACAGATACAGAACTTGT 49146
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
645 tagggacatactggagcaacttcgcacagcattgagaagtccttcgtccaaaca 704
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
49147 AGAGGACACTGGAAGAAATACAGAGATACATCAATAAATTTTAAACAAATCAAGTACA 49206
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
705 attacagcttgcagcttctcgtcgcgatctatctactcgcgatcagttacacga 764
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
49207 AATGAACCAAGATATCTTAGAGAACATATCTACTTCGCTCGTTCGAGTTACACCGA 49266
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
765 acaaatgcagtcataactacatgagaaattcctcagagagacataaagatgtgc 824
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
49267 GCAATATCGCAATTCCTTGAAGAGTTATCTCGAAGAAACAAAGATGAGGTAT 49326
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
825 agttattatgatagtgcttactccttcagcagatttgaagaatctgca 879
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
49327 AATAATTTCGAGTTCCATCCCGTTGATTGCTCTCAAGGTGATATGTGACA 49381
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 11  
 LMFCHR32\_17  
 WPCOMMENT

Sequence split into 28 fragments LOCUS LMFCHR32 Accession AL499622

Fragment Name	Begin	End	LOCUS	Accession
LMFCHR32_00	1	110000		
LMFCHR32_01	100001	210000		
LMFCHR32_02	200001	310000		
LMFCHR32_03	300001	410000		
LMFCHR32_04	400001	510000		
LMFCHR32_05	500001	610000		
LMFCHR32_06	600001	710000		
LMFCHR32_07	700001	810000		
LMFCHR32_08	800001	910000		
LMFCHR32_09	900001	1010000		
LMFCHR32_10	1000001	1110000		
LMFCHR32_11	1100001	1210000		
LMFCHR32_12	1200001	1310000		
LMFCHR32_13	1300001	1410000		
LMFCHR32_14	1400001	1510000		
LMFCHR32_15	1500001	1610000		
LMFCHR32_16	1600001	1710000		
LMFCHR32_17	1700001	1810000		
LMFCHR32_18	1800001	1910000		
LMFCHR32_19	1900001	2010000		
LMFCHR32_20	2000001	2110000		
LMFCHR32_21	2100001	2210000		
LMFCHR32_22	2200001	2310000		
LMFCHR32_23	2300001	2410000		
LMFCHR32_24	2400001	2510000		
LMFCHR32_25	2500001	2610000		
LMFCHR32_26	2600001	2710000		
LMFCHR32_27	2700001	2727709		

Continuation (18 of 28) of LMFCHR32 from base 1700001 (AL499622 Leishmania major chr

Query Match 5.9%; Score 87.2; DB 84; Length 110000;  
 Best Local Similarity 61.4%; Pred. No. 2e-10;  
 Matches 140; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

405 gaacacatactactcgtgtctgtgtgacctcaatgacatactgtgtgaggaattcgtg 464
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
32936 GACCACAGTACCAACTCTGTCGGGACGCTGACGCGTACTTGAGGTGGCGTCCGT 32995
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
465 caaagaagttactgagatggtggtgtccaggggtgtgtaaacctcaactgagattca 524
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
32996 TGCGGCGCTACCGAGATCATGTGGGCCCTTGCGCTTGGCAAGACACACGCTTGATGCA 33055
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
525 actagcaatcaatgatacaatcccaatccagtggaatgtgtgctgtgtggaagcaattta 584
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
33056 GCTAGCAGTTAGCTGTGCGCATGTCACATAGAGTTGCGAGGAGATGGGTGGCGCTTGT 33115
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
33116 CGTTGATTACGAGGACGACGCTTGTGCTGACGCGCTTAGAGCAAAATGCG 33163
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 12  
 AE001035/C  
 LOCUS  
 DEFINITION  
 AE001035 AE000782  
 ACCESSION  
 AE001035.1 GI:2689358  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Archaeoglobus fulgidus.  
 Archaeoglobus fulgidus  
 Archaea: Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 Archaeoglobus.

REFERENCE  
 1 (bases 1 to 10980)  
 Klenk,H.P., Clayton,R.A., Tomb,J., White,O., Nelson,K.E.,  
 Ketchum,K.A., Dodson,R.J., Gwinn,M., Hickey,E.K., Peterson,J.D.,  
 Richardson,D.L., Krelavage,A.R., Graham,D.E., Kyriades,N.C.,





```

gene
CDS
/feature="PEDIAVPGVEIRLLSCRCGACRTACPHGAVSGRIITIHREVDIRNTARLAAMDGI
RVFAKPGILEMMAHSI"
/feature="complement(1738. .3024)"
/feature="gene="MTH1380"
/feature="complement(1738. .3024)"
/feature="gene="MTH1380"
/feature="note="Function Code:5.02 - L-Amino Acid Metabolism,
glycine--serine and threonine metabolism ; similar to,
sp:LN:GLYA_METTH AC:P50436, p()=5.9E-203, pid=85%"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="AAB85857.1"
/db_xref="GI:2622490"
/translation="MEMIFVNSDYETEKIRQLMKDHSNMSSINLIASENITSSRY
KEALSDLSHRAEGLPGERLYEGRCYIDEIESEALVIELSKRLFAEHANVPTSGVVA
NACPFATAEVGDPIAMEVPGGHSISARAGVAGVFOIYHPPEENNDIADMD
KKITIEVKPRITILEGSLFLPHPEYBEALAEVYGARINTDCAHVIGLITAGTFOOP
LREGADMLVSTHKTFFPGGGIILCREELADIDEAVFGLVSNHHLHVAIGLGIAT
AEMLEFGAEYAOTINNARKLAENLHELGFVNLCEHIDFESHQVNDVSDIGRAEI
SKRLANNIILINKLLPMDVNRSDPDSGIRIGIQEITRRGKSESEVAYEYIKRVY
MDGKVRDEVAEFMSYTRVHYAFEDSEAYKYMELQ"
/feature="complement(3192. .5201)"
/feature="gene="MTH1381"
/feature="complement(3192. .5201)"
/feature="note="Function Code:10.08 - Metabolism of Macromolecules,
protein translation and modification ; similar to,
pir:LN:S48720 AC:S48720, p()=0, pid=88%"
/codon_start=1
/transl_table=11
/product="heterodisulfide reductase, subunit A"
/protein_id="AAB85858.1"
/db_xref="GI:2622491"
/translation="MLDTPMOGDLMAEKKEETMEPRIGVYVCHGNGIVGVIDEA
VRDAXKLPNVVSMDKYKSPDGOLEIKDKELGINRVYVACSPRIHEPFRFC
VEEGLNGLFEPANLPRDSDSWHMDDEGATEKAKLRYMAVAKALBLESKYS
VDKRLVIGGVAIGIOTRLDLDMGFTTYVEKRPISGRGQIDKTFPIIDCSMCLL
APKRVAVGKHNDIELITYAEVVDGYGNKVRKIEKPRYIDEDLCTGCGSCVECP
IEMPVFDEGIGKAVYIPPOAVPLCATIDKDYCIEMCLDEICERGAHVHOEPE
EIELEVGIIIVATGDAYADIPTEKLEYGYRHTNYTGLEERMINASGPTGKVIKFS
DGERKRAVAFHCGSRDEOIGKPCVSCVCCMIKMAKQIKDKMDPEVLYIMDR
AFGKEFEFRYSOEKYGKIFIRGPRAPALLENPDILTIVRESEDLILKRYEYDQMY
LGVLYVPEGSEKRLKOTIGLSKADGFLMEHPRKLPYDITLTDVYLAGAQQGKDI
DAVADASGAARAAIIPVYKGEVEIEPIVAVTSDVCGGCEVCIELCPFGAISTEEGHA
NNVALKCGCTCVAACPSGAMDQDHTEBOIMAOIRALMEQAK"
/feature="complement(5353. .6261)"
/feature="gene="MTH1382"
/feature="complement(5353. .6261)"
/feature="gene="MTH1382"
/feature="note="Function Code:14.01 - Unknown, Conserved protein;
similar to, pir:LN:F64428 AC:F64428, p()=3.7E-20, pid=24%"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AAB85859.1"
/db_xref="GI:2622492"
/translation="MIFWSTEVLSVLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLV
AIPSLIFSTIRADISGLDVLIPVICMIGTIGSTIGTAYIMARGLDGGKTMGLV
AAAMNSGELGYPVAGIFGSEGLVRAIFVDTGTTLMFTSGLILSHISGEGSRIMK
RAVFPPLMAFLVIRNMLPTGNGITIGTISGAAPVIMISLTLITNFRRLRS
VADATVSGRLILSPDMAGISVLAIRGILNFSVTLVLEASMPAMLAVALAENDID
VDLVSSCIFNSTILSVSLPLMSVLL"
/feature="complement(6276. .7211)"
/feature="gene="MTH1383"
/feature="complement(6276. .7211)"
/feature="gene="MTH1383"
/feature="note="Function Code:10.09 - Metabolism of Macromolecules,
DNA replication--modification--repair--and recombination ;
similar to, sp:LN:RADA_SULSO AC:Q55075, p()=6.6E-85,
pid=51%"
/codon_start=1
/transl_table=11
/feature="product="DNA repair protein Rada"
/feature="protein_id="AAB85860.1"
/feature="db_xref="GI:2622493"
/feature="translation="MVEIDLPVYGAKTAKOAKLRDAGFGGDMRLATATKELSVKAEIG
EGVAEYIEAARAEEKIDFETAPDVWERKKDGRITRTGSKALDELGGITETQATTEV
EGEGSGKSQLSHELAVTVQLEPERGGLDAEAVFIDETFTFPERIEQLANFELDLE
EVLNKHIAFAFSSHQILMAEKVNLIDGKNIRLIVYDLSLTAHRAEYVORELAT
ROOKLNDHLLTTONIANTYNAAVFYVNOVQARDAFGSPTRAGIGHVIGHAATYRIW
LRKGLAGKRIARLVDSPHLPEGECVFKITTAGIVD"
/feature="complement(7264. .7809)"
/feature="gene="MTH1384"
/feature="complement(7264. .7809)"
/feature="gene="MTH1384"
/feature="note="Function Code:10.09 - Metabolism of Macromolecules,
DNA replication--modification--repair--and recombination ;
similar to, pir:LN:F64444 AC:F64444, p()=0.0000042,
pid=21%"
/codon_start=1
/transl_table=11
/product="replication factor A related protein"
/protein_id="AAB85861.1"
/db_xref="GI:2622494"
/translation="MKLSIGNSRIETPASERDLGEPSPDEQEMHYPRDADDED
SRNVLLIEGLIEMSGRRILSIKPCSCNERLSDENICNFCGLDEPRYLLMTGRI
MDDTGEVMTTFEGREASELIEMTTDEVNINQSADESALBEEVBDLNGLTVRIGNA
DMDYSEELRFPKRVKKEL"
/feature="complement(7775. .9643)"
/feature="gene="MTH1385"
/feature="complement(7775. .9643)"
/feature="gene="MTH1385"
/feature="note="Function Code:10.09 - Metabolism of Macromolecules,
DNA replication--modification--repair--and recombination ;
similar to, pir:LN:F64444 AC:F64444, p()=3.7E-31, pid=45%"
/codon_start=1
/transl_table=11
/product="replication factor A related protein"
/protein_id="AAB85862.1"
/db_xref="GI:2622495"
/translation="MKELKREYERIKDRISPEFEELIEKKKEBIGDIFMDLTLIA
SYVVDILIEKNTMLSEKPEHNDITSLKEBAEFPVYGRVYISSPRTYRKREG
KLANVLIADDTGLRAVFTNENIKLKRFREDVIRIKVNRIGSGFERKEHLMPS
TYVLDVPEYPEREYREITTPGIDVEDDVNVARITGVSFEPEDGREGSPTS
LOIMDPTGSETTYLNNVDVILBELGKSGDAVKILMAOPRRDRKVTITRHSILRVY
PGATDYPERFEELVKTGDLHEKNVTVGLVTVKNDPVFEFRNDGTGVSKEIADV
TGSARVTLMEDEDRIRKINGDIIRISGANVEDDNOISYRINTNRTITLNPESDA
LKLVLVEYREOMKPMKISEILMEDEGEVDVVGRIFSLSDFEEREDGTGIVSME
LADETKIRISLMEDEKAERPMNIGDAVRIENRIRIGLTVSELSGRTRIYVLPED
MEDLPSFEELKMLYOTKRIADLEDDRIRIIRIARVYDLFEFRRODGDGPGIVRTA
EPADDTGSTRASIMDDAAEKPLSIDGPVYIENPRVYFRDMGGRLETOHRKQFD"
/feature="9947. .11206"
/feature="gene"
Query Match
Best Local Similarity 5.1%; Score 74.6; DB 1; Length 13968;
Matches 140; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
gene
412 atcaactggtcttgcgtgacccaatgacatactggtgcggaattcactgcaaga 471
|||||
6977 ATCAACAAGGGAGACAGCCCTGATGATAGTACGCGAATCGACACAGCA 6918
|||||
472 gtactgaagatcggtgcgcgccaggggtgtgcaaaactaactggggtatcaactaga 531
|||||
6917 ATAACCCAGGTTTTCGAGAAATTCGGGTCAAGTCAAGCTGTCCTCATGAACTTGA 6858
|||||
532 atcaatgacaatcccaatggaatggtggtgccttggggaagcaattatagaat 591
|||||
6957 GTCAAGGTTTCAAGCTGCCAGAGAGAGAGGAGACTGATGCTGTTTTCATAGAC 6798
|||||
592 acagaagggcattcatggttgaactgtctaccagaattgctaaggggtatagaagac 651
|||||
6797 ACAAGAGAACATATCTCTCTGAGAGGATTCAGACATATGCAAAATCCTTTGAACTTGAC 6738
|||||
652 atactggag 660
|||||

```

Db 6737 CTGAGGAG 6729

RESULT 14  
 SSU45310 1699 bp DNA BCT 18-JUN-1996  
 LOCUS Sulfolobus solfataricus rada gene, complete cds.  
 DEFINITION  
 ACCESSION U45310  
 VERSION U45310.1 GI:1378035  
 KEYWORDS  
 SOURCE Sulfolobus solfataricus.  
 ORGANISM Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.

REFERENCE 1 (bases 1 to 1699)  
 Sandler, S.J., Saito, L.H., Samra, H.S. and Clark, A.J.  
 Archaeal-like genes from three archaeal species with putative protein products similar to Rad51 and Dmc1 proteins of the yeast

REFERENCE 2 (bases 1 to 1699)  
 Sandler, S.J., Saito, L.H., Samra, H.S. and Clark, A.J.  
 Direct Submission  
 Submitted (11-JUN-1996) Steven J. Sandler, Molecular and Cell Biology, University of California, 401 Barker Hall #3202, Berkeley, CA 94720-3202, USA

JOURNAL MEDLINE  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

FEATURES  
 source location/Qualifiers

1..1699  
 /organism="Sulfolobus solfataricus"  
 /db\_xref="taxon:2287"  
 352..1326  
 /gene="rada"  
 352..1326  
 /gene="rada"  
 /note="RecA-Rad51 homolog"  
 /codon\_start=1  
 /transl\_table=1  
 /product="rada"  
 /protein\_id="AAC44123.1"  
 /db\_xref="GI:1378036"

BASE COUNT 615 a 223 c 354 g 507 t  
 ORIGIN

Query Match 4.9%; Score 72; DB 3; Length 1699;

Best Local Similarity 52.3%; Pred. No. 7.7e-07;  
 Matches 159; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

334 aatgacccaagaagaagcttcaagagcccaagatcgctggatattcctcgtat 393  
 538 AAGAAGCGAGAGCGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 597  
 394 gagctgtcaagaagaacacatcactactgttctggtgacctaatgacatactgtgac 453  
 598 AGAATGAACCTTAAGAAGATATCTACTGCTGATCAAGCACTAGATGCTTACTTAC 657  
 454 gggattcaatcgaagaagttactgagatggtgtgctgcccaagggtgtgtaaacctaa 513  
 658 GGTATGAACCTTAAGAAGATATCTACTGCTGATCAAGCACTAGATGCTTACTTAC 717  
 514 ctgggattcaatcgaagaagttactgagatggtgtgctgcccaagggtgtgtaaacctaa 573  
 718 CTATGATCATCACTTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 777  
 574 aaagcattatagatagacagggcagtttctggttgaagctgtctacagattctc 633  
 778 AAGCAGTTTCACTTGAATACAGAAAGTACTTTCAGATGGAAGAAATGAATATAGGCT 837

OY 634 gaag 637  
 Db 838 AAGG 841

RESULT 15  
 AP000058 202000 bp DNA BCT 06-APR-2000  
 LOCUS Aeropyrum pernix genomic DNA, section 1/7.  
 DEFINITION  
 ACCESSION AP000058 BA000002  
 VERSION AP000058.1 GI:5103388  
 KEYWORDS  
 SOURCE Aeropyrum pernix (strain:K1) DNA.  
 ORGANISM Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;

REFERENCE 1 (sites)  
 Kawarabayashi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y.,  
 Jin, no, K., Takahashi, M., Sekine, M., Baba, S., Anka, A., Kosugi, H.,  
 Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Nakazawa, H.,  
 Takamizawa, M., Masuda, S., Funahashi, T., Tanaka, T., Kudoh, Y.,  
 Yamazaki, J., Kusuda, N., Oguchi, A., Aoki, K., Kubota, K.,  
 Nakamura, Y., Nomura, N., Sako, Y. and Kikuchi, H.

Complete genome sequence of an aerobic hyper-thermophilic  
 crenarchaeon, Aeropyrum pernix K1  
 DNA Res. 6 (2), 83-101 (1999)

JOURNAL MEDLINE  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT

2 (bases 1 to 202000)  
 Tanaka, T., Hino, Y., Kawarabayashi, Y. and Kikuchi, H.  
 Direct Submission  
 Submitted (14-DEC-1998) to the DDBJ/EMBL/Genbank databases. Yutaka  
 Kawarabayashi, National Institute of Technology and Evaluation,  
 Biotechnology Center; Nishihara 2-49-10, Shibuya-ku, Tokyo  
 151-0066, Japan (E-mail:kyutak@kazusa.or.jp, Tel: +81-3-3481-8951,  
 Fax: +81-3-3481-8424)  
 Kawarabayashi, Y. is officially affiliated with the National  
 Institute of Bioscience and Human Technology, Tsukuba, Ibaraki  
 305-0046, Japan.  
 Nomura, N. and Sako, Y. are at Kyoto University, Sakyo-ku, Kyoto  
 606-8502, Japan  
 The other authors are at the National Institute of Technology and  
 Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.  
 All the sequence with length 100 codons or more between ATG or GTG  
 and stop codon are defined as CDS  
 Homology analysis is performed by Smith-Waterman algorithm against  
 Genbank and Genpept release 109; EMBL release 56.0; Swissprot  
 release 36.0; PIR-Protein release 57.0; and OML release 31.0.  
 E-mail address for comments and questions: genome@pentle.go.jp  
 Restriction map, ORF organization, sequence alignment and more  
 information are available at W.W.W. site of Biotechnology Center,  
 URL: <http://www.mld.nite.go.jp/>.

FEATURES  
 source location/Qualifiers

1..202000  
 /organism="Aeropyrum pernix"  
 /strain="K1"  
 /db\_xref="taxon:56636"  
 complement(213..938)  
 /gene="APE0001"  
 complement(213..938)  
 /gene="APE0001"  
 /codon\_start=1  
 /transl\_table=11  
 /product="241aa long hypothetical protein"  
 /protein\_id="BA00001.1"  
 /db\_xref="GI:5103389"  
 /translation="WVDIISLLSLSPGVIGFLIVSPGSIWTPVKEISIGVYVSR  
 VYVAKSLGLSLTLASLISFVGAAYGTTQASLTALLALLIVYVYSRMLAEIE  
 SINQVLEGFEEVSGIKKLYTILLVLVISIVSISGLKISGAGTILASHLSE  
 ILAAGTVLFVYVRLINKEGKY"  
 complement(938..1276)  
 /gene="APE0002"



Db 89267 CAGGAACCTTGACGAGCTGCTGGCGGCTATATAGACTAAGACTATATAACGAGCTTT 89326  
Qy 486 tggcgtcccaagggttggtaaaactcaactgggattcaactagcaatcaatgtaaat 545  
Db 89327 CGGTGAGTTGCGGAGCGGCAGACTCAGATATGCCATCAGCTATCTGTTAACGTCAGCT 89386  
Qy 546 cccagtggaaatgtggtggccttggtagaaagcagttatatagatacagaggcaatt 605  
Db 89387 CCCCAGAGACAGAGGAGGCTTCGAGGGTTAAGGCGGTATACGTAGATACCGAGGCACTT 89446  
Qy 606 catggttgaacgtgtctaccagaattgtgaag 638  
Db 89447 CAGGTGGAGAGGATAGAGCAGATGGCTGCTGG 89479

Search completed: October 2, 2001, 05:54:16  
Job time: 3732 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2001, 16:27:27 ; Search time: 11.19 Seconds

(Without alignments)  
900.010 Million cell updates/sec

Title: US-09-537-654-2

Perfect score: 1530

Sequence: 1 MGDQSGSRNGPQOKYVSGAQ.....VTGKGRDVAASNHRARVT 294

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	576.5	37.7	376	1	R51C_HUMAN
2	387.5	25.3	324	1	RADA_SUIISO
3	354.5	23.0	337	1	RADA_ARCVU
4	352.5	23.0	311	1	RADA_METH
5	337	22.0	334	1	DMC1_YEAST
6	323	21.1	350	1	R51B_HUMAN
7	314.5	20.6	340	1	DMC1_MOUSE
8	313	20.5	340	1	DMC1_HUMAN
9	313	20.5	349	1	DMC1_LILLO
10	313	20.5	350	1	R51B_MOUSE
11	312	20.4	345	1	DMC1_SOYBN
12	312	20.4	352	1	RADA_METJA
13	304	19.9	324	1	DLH1_CANAL
14	299.5	19.6	339	1	RAS1_USRMA
15	297.5	19.4	332	1	DMC1_SCHPO
16	290	19.0	342	1	DMC1_ARATH
17	287.5	18.8	339	1	RAS1_CHICK
18	287.5	18.8	343	1	RADA_HALVO
19	284.5	18.6	339	1	RAS1_CRGR
20	282.5	18.5	336	1	R511_XENLA
21	281.5	18.4	339	1	RAS1_HUMAN
22	280.5	18.3	336	1	R512_XENLA
23	280.5	18.3	339	1	RAS1_MOUSE
24	277.5	18.1	400	1	RAS1_YEAST
25	276.5	18.1	339	1	RAS1_RABIT
26	276	18.0	365	1	RAS1_SCHPO
27	275.5	18.0	336	1	RAS1_DROME
28	272.5	17.8	342	1	RAS1_LICES
29	241.5	15.8	460	1	RAS1_YEAST
30	217.5	14.2	328	1	R51D_HUMAN
31	209	13.7	329	1	R51D_MOUSE
32	206	13.5	221	1	R51L_ARCFU
33	199	13.0	212	1	R51L_METJA

34	199	13.0	228	1	R51L_PYROH	057859 pyrococcus
35	189	12.4	234	1	R51L_METTH	027728 methanobact
36	174.5	11.4	216	1	R51L_METTH	050248 methanococc
37	164.5	10.8	350	1	RH55_SCHPO	014129 schizosacch
38	134.5	8.8	452	1	RADA_LISMO	048761 listeria mo
39	133	8.7	458	1	RADA_BACSU	037572 bacillus su
40	127.5	8.3	406	1	RAS5_YEAST	P38953 saccharomyc
41	125.5	8.2	458	1	RADA_HAEIN	P45266 haemophilus
42	118	7.7	343	1	RECA_XANCP	060101 xanthomonas
43	117.5	7.7	448	1	RADA_RICPR	092d04 rickettsia
44	117.5	7.7	453	1	RADA_PSEAE	P96963 pseudomonas
45	117.5	7.7	480	1	RADA_MYCTU	053570 mycobacteri

## ALIGNMENTS

RESULT	1	STANDARD	PRT	376 AA.
ID	R51C_HUMAN			
AC	043502;			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DE	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	DNA REPAIR PROTEIN RAD51 HOMOLOG 3.			
GN	RAD51L2 OR RAD51C.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-98136197; PubMed=9469824;			
RA	Dosenh M.K., Collins D.W., Fan W., Lennon G.G., Albalade J.S., Shen Z.,			
RA	Schild D.;			
RT	Isolation and characterization of RAD51C, a new human member of the			
RT	RAD51 family of related genes.*			
RL	Nucleic Acids Res. 26:1179-1184(1998).			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).			
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TISSUES, WITH			
CC	HIGHEST EXPRESSION IN TESTIS, HEART MUSCLE, SPLEEN AND PROSTATE.			
CC	-1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO			
CC	PROKARYOTIC RECA PROTEIN.			
CC	-1- This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
DR	EMBL; AF029669; AAC39604.1; -			
DR	MIM; 602774; -			
DR	PROSITE; PS50162; RECA_2; 1.			
KW	DNA-binding; ATP-binding; Nuclear protein.			
FT	NP_BIND 125 132 ATP (POTENTIAL).			
SQ	SEQUENCE 376 AA; 42189 MW; 3AAAD3C1C0851E0 CRC64;			

Query Match 37.7%; Score 576.5; DB 1; Length 376;  
Best local Similarity 43.9%; Pred. No. 2.2e-42;  
Matches 125; Conservative 47; Mismatches 98; Indels 15; Gaps 4;

QY	22	AWMEFDELISOKHTITGSGDLNIIIGGCHKEVEIGSGPVGKTQGIOLAIWQIPV	81
DB	87	ALLELQEHRTGQGITTFGCALDDILGGVPLMKTTETICAPGVGTQLCMQALVDVQD-	145
QY	82	EC-GGAGRAVYIDEGSPMERVVOIACGCTRDI-LEHFPSSHRSKSSVQKQLQPERFL	139
DB	146	ECGGVAGAEVFIIDEGSPMDVRVDVLAFCIQHQLIAEKHKGSEHRALEDFLNTL	205
QY	140	ADIIYRISYIEQIAVINIMKEFLREHKDVRIIIVDSVTFHFRODFEDLALRTVLSGL	199

```

Db      206  SHIIYFRCKDYTELLAQYAILPDLFSEHSKVALIVDQIAFPERHDDLSTRTLLNGL 265
Oy      200  SLKLMKIATYVLAAYVLLNQVTTKTEGSGFOLTALGDSMSHCNRRLLHNGMDERAH 259
Db      266  AQOMSTSLNNHRLAVLTNQMTKTKIDRNQALLVPAALGSGWGHAAATRIILFHWDRQRLAT 325
Oy      260  LDKSSLPASAPAVTQKGRDAV-----SSNHHKRR 292
Db      326  LYKSSQKCEYVLFQIKRQGRFDYVTSACSQTQTESLSTRKRSR 370

RESULT      2
RADA_SULSO
ID      RADA_SULSO      STANDARD;      PRT;      324 AA.
AC      055075;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      15-DEC-1998 (Rel. 37, Last annotation update)
DE      DNA REPAIR PROTEIN RADA.
GN      RADA.
OS      Sulfolobus solfataricus.
OX      Archaeae; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
XX      NCBI_TaxID=2287;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      MEDLINE=96251095; PubMed=8668545;
RZ      Sandler S.J., Satlin L.H., Samra H.S., Clark A.J.;
RT      "reca-like genes from three archaeal species with putative protein
RT      products similar to Rad51 and Dmc1 proteins of the yeast
RT      Saccharomyces cerevisiae." 24:2125-2132(1996).
RL      Nucleic Acids Res. 24:2125-2132(1996).
CC      -I- FUNCTION: PROBABLY INVOLVED IN DNA REPAIR AND IN HOMOLOGOUS
CC      RECOMBINATION AND REQUIRED FOR MEIOTIC RECOMBINATION.
CC      -I- SIMILARITY: STRONG TO EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
CC      PROKARYOTIC RECA PROTEIN.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
CC      EMBL; U45310; AAC4123.1; -.
DR      HSSP; P03017; 2RBB.
DR      PROSITE; PS50162; RECA_2; 1.
DR      PROSITE; PS50163; RECA_3; 1.
KW      DNA damage; DNA recombination; ATP-binding; DNA-binding.
FT      NP_BIND 114 121
FT      SEQUENCE 324 AA; 35866 MW; B08CA699AA40A88 CRC64;

Query Match      25.3%; Score 387.5; DB 1; Length 324;
Best Local Similarity 34.2%; Pred. No. 3.5e-26;
Matches 97; Conservative 61; Mismatches 93; Indels 33; Gaps 5;

Oy      13  QKYYSAQNAAMP-----MFSDLSQCHNITGSSDDLIIIGGTHCKEYVEIGSPV 64
Db      59  QIIKEARADLDIRFTALEVKKERNNYKISTGSQALDGLAGLETFTMEFGEFSGS 118
Oy      65  GKTGLOIALINVOJPEVCGGGLGKAVYIDTSGSPFVEVVOIAEGCIDIIEHPHSH 124
Db      119  GKTOLCHQSLVNVQLPRENGGLSGRAVYIDTSGTFRMEIEMAKALGDI----- 169
Oy      125  KSSSVQKQLQPERFLADYIFRICSTEQIAVINMTEKFLREHKDURIYIDSVTFH 184
Db      170  -----DNVMNNIYIRAIINTDHOIAIVDDLOELYSKDPISIKLIYVDSVSH 218
Oy      185  DE---EDLALRRRVISGLSKLMIKAKTYNLAAYVLLNQVTTKFTBESFQDLTAL- 240
Db      219  EPPGRRIIVLRQOKLMIKHOLTRIALEVYDVIITNQVMARPDMEFYDPTVAVGSH 278

```

Oy	24.1	HSCRTLLHMN-GENERYALDKSPSLVAPSVATGCGJDA	283
Dd	27.9	HVPGRIOIOLKSRGNRRRIARVVDAPHLEGEVVFALTEGIRDA	322
<hr/>			
RESULT 3			
ID	RADA_ARCFU	STANDARD:	PRT; 337 AA.
DC	O29269;		
DT	15-DEC-1998 (Rel. 37, Created)		
DM	15-DEC-1998 (Rel. 37, Last sequence update)		
DN	15-DEC-1998 (Rel. 37, Last annotation update)		
DE	DNA REPAIR PROTEIN RADA.		
DD	RADA OR AF0993.		
OS	Archaeoglobus fulgidus.		
OC	Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;		
CC	Archaeoglobus.		
XX	NCHI_TaxID=2234;		
RM	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-VG-16 / DSM 4304 / ATCC 49558;		
RX	MEDLINE=96049343; PubMed=9389475;		
RA	Kleck H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,		
RA	Rechun K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,		
RA	Richardson D.L., Kerlavage A.R., Graham D.E., Kyteides N.C.,		
RA	Fischmann R.D., Quackenbush J., Lee N.H., Strydom G.G., Gill S.,		
RA	Kirknes E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,		
RA	Peterson S., Kelch C.T., McNeill L.K., Badger J.H., Glodet A., Zhou L.,		
RA	Overbeek R., Goeyne J.D., Weidman J.F., McDonald L., Uitterback T.,		
RA	Cotton M.D., Spriggs T., Artlach P., Kaine B.P., Sykes S.M.,		
RA	Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,		
RA	Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,		
RA	Venter J.C.;		
RT	"The complete genome sequence of the hyperthermophilic, sulphate-		
RT	reducing archaeon Archaeoglobus fulgidus.";		
RL	Nature 390:364-370(1997).		
CC	-I- FUNCTION: PROBABLY INVOLVED IN DNA REPAIR AND IN HOMOLOGOUS		
CC	RECOMBINATION AND REQUIRED FOR MEIOTIC RECOMBINATION.		
CC	-I- SIMILARITY: STRONG TO EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO		
CC	PROKARYOTIC RECA PROTEIN.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outpostation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@sib-sib.ch).		
CC	-----		
DR	EMBL; AE001035; AAB90248.1; "		
DR	TIGR; AF0993; "		
DR	PROSITE; PS50162; RECA_2; 1.		
DR	PROSITE; PS50163; RECA_3; 1.		
DR	DNA damage; DNA recombination; ATP-binding; DNA-binding.		
KW	NP_BIND 115 122 ATP (POTENTIAL).		
FT	SEQUENCE 337 AA; 37162 MW; 7688592BAA3BDFA3 CRC64;		
SQ			
<hr/>			
Query Match 23.2%; Score 354.5; DB 1: Length 337;			
Best Local Similarity 32.4%; Pred. No. 2.6e-23;			
Matches 95; Conservative 50; Mismatches 105; Indels 43; Gaps 6;			
Oy	16	VSGANAMDMFDEFDELISGHITTTSGDLDNLIELGGHCHKEVTETIGGVPGVGKTGLQLAI	75
Dd	71	IGGFESGCKVLERRRSYVKITTGSKDLDBELAGCVGTQAITEFPFEFGSGKRQICHOLAV	130
Oy	76	NVOIFVECGGGLGAKVAVIDTESGFWERYRYOIAECIRDILEHPHSHEKSSSVOKLOLP	135
Dd	131	NVOLDEDEGEGSLGYIIIDTFENTFRPERIIOAEEA-----KGIDG	170
Oy	136	EERFLADYYFRICSTEQALVIN----YMEKRLREHKOVRIYIDSYTFHRQDF---ED	188

Db 171 NEVLANIYAAQYNSNMQLVNDNAKELAKKGRPVRLITVDSLSMHPFAEYVGRGT 230  
 QY 189 LALFRFVLSGLSLKMLKIKATYNLAIVLNOYTKFTBEGSFQTLALGDSW-----S 240  
 Db 231 LADROCKLRHLHDLKMFSELYNAAIVTNQVMAR-----PDVLFSDPTKPVGCHIVA 283  
 QY 241 HSCNTNRLILHMNGNE-RVAHLDKSPSLPVASAPYAVTGKIGINDAVSNHKKRAR 292  
 Db 284 HTATFRYILKKGKDLRIARLIDSPHLEGEAIFVTEGIDAEKDKKKKK 336

## RESULT 4

RADA\_METTH

ID RADA\_METTH STANDARD: PRT: 311 AA.

AC 027436;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE RADA PROTEIN.

GN RADA OR MTH1383.

OS Methanobacterium thermoautotrophicum.

OC Archaea: Euryarchaeota; Methanobacteriales; Methanobacteriaceae;

OC Methanobacter.

OX NCBI\_TaxID=145262;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-DELTA H;

RX MEDLINE=96037514; PubMed=9371463;

RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,

RA Aldredge T., Bashirzaden R., Blakely D., Cook R., Gilbert K.,

RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,

RA Spadafora R., Viare R., Wang Y., Wierzbowski J., Gibson R.,

RA Jivani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,

RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,

RA Daniels C.J., Mao J.-I., Rice P., Nollan J., Reeve J.N.;

RT "Complete genome sequence of Methanobacterium thermoautotrophicum

RT delah: functional analysis and comparative genomics.";

RL J. Bacteriol. 179:7135-7155(1997).

CC -I- FUNCTION: PROBABLY INVOLVED IN DNA REPAIR AND IN HOMOLOGOUS

CC RECOMBINATION AND REQUIRED FOR MEIOTIC RECOMBINATION.

CC -I- SIMILARITY: STRONG TO EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO

CC PROKARYOTIC RECA PROTEIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC DR EMBL: AE000901; AA85860.1; -

CC PROSITE: PS50162; RECA\_2; 1.

CC PROSITE: PS50163; RECA\_3; 1.

CC DR PROSITE: PS50163; RECA\_3; 1.

CC NP\_BIND: DNA recombination; ATP-binding; DNA-binding.

CC FT NP\_BIND 104 111 ATP (POTENTIAL).

CC SEQUENCE 311 AA; 34125 MW; 3A0DF609A8785DB CRC64;

CC -----

CC Query Match 23.0%; Score 352.5; DB 1; Length 311;

CC Best Local Similarity 33.6%; Pred. No. 3.5e-23;

CC Matches 96; Conservative 46; Mismatches 111; Indels 33; Gaps 5;

CC -----

CC 10 GPOQKYVSGA-----QNAWMFSDLSQKHITTSQNDILGSGIHCKEYTEIGV 61

CC 46 GVAEYVIEAARAETIDEIADVEMERKKDVGRTTSGKALDELGGIETQATTEVGE 105

CC 62 PEGVKTQGLDGLAINVQVPEEGGAGKAYIDEGSFMEVRYQIAEGCTRIDLIEHPH 121

CC 106 FGSGKSQSLSHLAVTVQPEERGGDAEAVFIDENFRPRRIQAINAF----- 155

CC 122 SHEKSSVQKQLOPRFLADIIYFRICSTQIAIVINMEKFLREBKDVRIIVISVTFH 181

Db 156 -----ELDLEEVINKIHIAFPNSSHQLMAEKVNELIOEGKNIRLIVDSLTAH 205  
 QY 182 PROPE-----EDLALFRVLSGLSLKMLKIKATYNLAIVLNOYTKFTBEGSFQTLALGDS 238  
 Db 206 FRAYVGREALATQOKLNOHLHTLQNIANTYNAAYVTNQVAPDPAGFSGPTKAIGH 265  
 QY 239 -WSISCTNRLILHMN-ENERYAHLDKSPSLPVASAPYAVTGKIGIRD 282  
 Db 266 VLGHAATYRIWKLKGLAGKRIARLIVDSPLPEGEVCKITTAGIVD 311

## RESULT 5

DMC1\_YEAST

ID DMC1\_YEAST STANDARD: PRT: 334 AA.

AC P25453;

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE MEIOTIC RECOMBINATION PROTEIN DMC1.

GN DMC1 OR ISC2 OR YER179W.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SK1;

RX MEDLINE=92257586; PubMed=1581960;

RA Bishop D.K., Park D., Xu L., Kleckner N.;

RA [DMC1: a meiosis-specific yeast homolog of E. coli recA required for

RA recombination, synaptonemal complex formation, and cell cycle

RA progression.";

RL Cell 69:439-456(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-SK1;

RX MEDLINE=93204898; PubMed=8455558;

RA Kobayashi T., Hotta Y., Tabata S.;

RA "Isolation and characterization of a yeast gene that is homologous

RA with a meiosis-specific cDNA from a plant.";

RL Mol. Gen. Genet. 237:225-232(1993).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-SK1;

RX MEDLINE=93204898; PubMed=8455558;

RA Kobayashi T., Hotta Y., Tabata S.;

RA "Isolation and characterization of a yeast gene that is homologous

RA with a meiosis-specific cDNA from a plant.";

RL Mol. Gen. Genet. 237:225-232(1993).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN-SK1;

RX MEDLINE=93204898; PubMed=8455558;

RA Kobayashi T., Hotta Y., Tabata S.;

RA "Isolation and characterization of a yeast gene that is homologous

RA with a meiosis-specific cDNA from a plant.";

RL Mol. Gen. Genet. 237:225-232(1993).

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN-SK1;

RX MEDLINE=93204898; PubMed=8455558;

RA Kobayashi T., Hotta Y., Tabata S.;

RA "Isolation and characterization of a yeast gene that is homologous

RA with a meiosis-specific cDNA from a plant.";

RL Mol. Gen. Genet. 237:225-232(1993).

RN [6]

RP SEQUENCE FROM N.A.

RC STRAIN-SK1;

RX MEDLINE=93204898; PubMed=8455558;

RA Kobayashi T., Hotta Y., Tabata S.;

RA "Isolation and characterization of a yeast gene that is homologous

Query Match	22.0%	Score 337	DB 1	Length 334
Best Local Similarity	33.5%	Pred. No. 8,2e-22		
Matches 87	Conservative 46	Mismatches 93	Indels 34	Gaps 6
DR EMBL: M87549; AAA34571.1; -				
DR EMBL: D10865; BAA01637.1; -				
DR EMBL: U18922; AAB64706.1; -				
DR PIR: S30857; S30857.				
DR PIR: S30276; S30276.				
DR PIR: A38214; A38214.				
DR SGD: S0000981; DMC1.				
DR PROSITE: PS50162; RECA.2; 1.				
DR PROSITE: PS50163; RECA.3; 1.				
DR Scourlation: Meiosis: Cell cycle: ATP-binding; Nuclear protein.				
FT NP_BIND 121 128 ATP (POTENTIAL).				
SO SEQUENCE 334 AA; 36612 MW; 00B0303D7FAF5CB3 CRC64;				
Query 35	ITTTGGDLNDILGGGICKEVEYIGGVPVGKQTGLQIALINVOIPVECGGLGGKAYID 94			
Db 96	LSTSGKQLSDILGGGIMTMSITEVEFGEEFCRGKTQMSHLCVTTOLPREMGGEKGAYID 155			
Qy 95	TESEFWERYXVYLAEGCINDILEFPHSHKSSVQKLOPERFLADYYRIRCSYEOI 154			
Db 156	TEGFPERPKQIAEG-----YELDEPESCLANYSIKRALNSEHOM 195			
Qy 155	AVIYMKERFLREKDVRIYIDSVTFHFROD---EDIALPTRLVLSGLSKIMKIATYTN 211			
Db 196	ELVBOQLGEEL-SSQDYRLIYVDISIMANFVYDYGCGELSEROOKINQHLFLNLAEFFN 254			
Qy 212	LAVVLLNQVYTKTTEGSPFOLTAL-----GDSMSHSCITRLIILH-WNGERVAHDKS 263			
Db 255	VAVLVLTNVOVS--DPGSALEFASADGRKPIGSHVILAHSAFRIILKRGDERVAKLQDS 312			
Qy 264	PSLPVASAPYAVTGTGKIRDA 283			
Db 313	PDMEPKCEGVYVIGEKGITDS 332			
RESULT 6				
R51B_HUMAN	STANDARD;	PRT;	350 AA.	
AC 015315;	060914;			
DT 15-DEC-1998	(Rel. 37, Created)			
DT 15-DEC-1998	(Rel. 37, Last sequence update)			
DT 01-OCT-2000	(Rel. 40, Last annotation update)			
DE	DNA REPAIR PROTEIN RAD51 HOMOLOG 2 (R51H2).			
GN	RAD51L1 OR RAD51B OR REC2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	SEQUENCE FROM N.A. (1ST FORM).			
RP	MEDLINE=97352813; PubMed=9207106;			
RA	Rice M.C., Smith S.T., Bullrich F., Havre P., Kmiec E.B.;			
RT	"Isolation of human and mouse genes based on homology to REC2, a			
RT	recombinational repair gene from the fungus <i>Ustilago maydis</i> ."			
RL	Proc. Natl. Acad. Sci. U.S.A. 94:7417-7422(1997).			
RN	SEQUENCE FROM N.A. (1ST FORM).			
RP	MEDLINE=98110585; PubMed=9441753;			
RA	Albala J.S., Thelen M.P., Prange C.K., Fan W., Christensen M.,			
RA	Thompson L.H., Lennon G.G.;			
RT	"Identification of a novel human RAD51 homolog, RAD51B."			
RL	Genomics 46:476-479(1997).			
RN	SEQUENCE FROM N.A. (2ND FORM).			
RP	MEDLINE=98181067; PubMed=9512535;			
RA	Cattwight R., Dunn A.M., Simpson P.J., Tamblini C.E., Thacker J.;			
RT	"Isolation of novel human and mouse genes of the recA/RAD51			
RT	recombination repair gene family."			
RL	Nucleic Acids Res. 26:1653-1659(1998).			

CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;  
CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN A WIDE RANGE OF TISSUES.  
CC -1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO  
CC PROKARYOTIC RECA PROTEIN.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: U92074; AAB63358.1; -;  
DR EMBL: U84138; AAC39723.1; -;  
DR EMBL: Y15571; CAA75680.1; -;  
DR MIM: 602948; -;  
DR PROSITE: PS50162; RECA\_2; 1.  
KW DNM-binding; ATP-binding; Nuclear protein; Alternative splicing.  
FT NP\_BIND 108 115 ATP (POTENTIAL).  
FT VARSPIC 346 350 AYGNS -> GGEKP (IN ISOFORM 2).  
SQ SEQUENCE 350 AA; 38257 MW; CE9A92DDC394EF3B0 CRC64;

[illegible]

```

RP SEQUENCE FROM N.A.
RC STRAIN-129/SV, TISSUE-Testis;
RX MEDLINE-96173646; PubMed-8602360;
RA Habu T., Taki T., West A., Nishimune Y., Morita T.;
RT "The mouse and human homologs of DMCL, the yeast meiosis-specific
RT homologous recombination gene, have a common unique form of exon-
RT skipped transcript in meiosis."
RL Nucleic Acids Res. 24:470-477(1996).
CC -1- FUNCTION: MAY PARTICIPATE IN MEIOTIC RECOMBINATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- TISSUE SPECIFICITY: TESTIS.
CC -1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
CC PROKARYOTIC RECA PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D58419; BAA09590.1; -
DR EMBL: D64108; BAA10969.1; -
DR HSSP: P03017; 2REB.
DR MGD: MGI:105393; Dmclh.
DR PROSITE: PS50162; RECA_2; 1.
DR PROSITE: PS50163; RECA_3; 1.
DR Meiosis; Cell cycle; ATP-binding; Nuclear protein.
KW NP_BIND 126 133 ATP (POTENTIAL).
FT SEQUENCE 340 AA; 37821 MW; 23EAB08D6E68637 CRC64;
SQ
Query Match 20.6%; Score 314.5; DB 1; Length 340;
Best Local Similarity 33.2%; Pred. No. 7.3e-20;
Matches 87; Conservative 43; Mismatches 97; Indels 35; Gaps 7;
QY 34 HITGGGDLNDILGGGHCKEVTEIGVGVGKTQIGIOLAINVOIPVECGGLGKAVYI 93
DB 100 HITGSGQFEDKLTGGGIESMAITEAFGEFRTQTQLSHLTCVTAQLPGTGTSGGKIIFI 159
QY 94 DTGSGFMERYVQIAEGCIRDLIEHPHSHKSSSVQKQLOPERFLADIIYFRICSTEQ 153
DB 160 DTEMFPRDR-----LRDIADRFNVDR-----EAVLDNVYARAYTSEHQ 199
QY 154 IAVIYVM-EKFLREHKDRIYIIVDFHFRDFE--DLAIFRIVSGSLKIMTKAKT 209
DB 200 MELLDYVAAKHEHAGIFKLLIIDSIMALFRVDSGKGLAEKQKLAOMLSRLQKISEE 259
QY 210 YNLAVVLLNQVTFKTEGSGFQLTLAL-----GDSWSHSCNRLILH-WNGNERVAHLD 261
DB 260 YNVAVFVNTQMT---ADPGATMTFOADPKPKPIGHIILAHSTRTISLRKGGELRIAKIY 316
QY 262 KPSPLPVAAPYAVTGGKIRDA 283
DB 317 DSPENPENEAFTTAGIGIDA 338

```

```

RESULT 8
DMCL_HUMAN STANDARD; PRT; 340 AA.
AC Q14565; Q99498;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MEIOTIC RECOMBINATION PROTEIN DMCL/LIM15 HOMOLOG.
GN DMCL OR DMC1H OR LIM15.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE-Testis;
RX MEDLINE-96127532; PubMed-8590282;
RA Sato S., Seki N., Hotta Y., Tabeta S.;
RT "Expression profiles of a human gene identified as a structural
RT homologue of meiosis-specific recA-like genes."
RL DNA Res. 2:183-185(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Testis;
RX MEDLINE-96173646; PubMed-8602360;
RA Habu T., Taki T., West A., Nishimune Y., Morita T.;
RT "The mouse and human homologs of DMCL, the yeast meiosis-specific
RT homologous recombination gene, have a common unique form of exon-
RT skipped transcript in meiosis."
RL Nucleic Acids Res. 24:470-477(1996).
CC -1- FUNCTION: MAY PARTICIPATE IN MEIOTIC RECOMBINATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
CC PROKARYOTIC RECA PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D63882; BAA09932.1; -
DR EMBL: D64108; BAA10970.1; -
DR HSSP: P03017; 2REB.
DR MIM: 602721; -.
DR PROSITE: PS50162; RECA_2; 1.
DR PROSITE: PS50163; RECA_3; 1.
DR Meiosis; Cell cycle; ATP-binding; Nuclear protein.
KW NP_BIND 126 133 ATP (POTENTIAL).
FT CONFLICT 37 37 I -> N (IN REF. 2).
FT CONFLICT 183 183 P -> A (IN REF. 2).
FT SEQUENCE 340 AA; 37707 MW; EBBBDE57EB440402 CRC64;
SQ
Query Match 20.5%; Score 313; DB 1; Length 340;
Best Local Similarity 32.7%; Pred. No. 9.8e-20;
Matches 89; Conservative 45; Mismatches 98; Indels 40; Gaps 8;
QY 29 ELGK-----HITGGGDLNDILGGGHCKEVTEIGVGVGKTQIGIOLAINVOIPVE 83
DB 90 EYSEKRRWFHITGSGQFEDKLTGGGIESMAITEAFGEFRTQTQLSHLTCVTAQLPGAG 149
QY 84 GGLGKAVYIDTSGFMERYVQIAGCIRDLIEHPHSHKSSSVQKQLOPERFLADIIY 143
DB 150 GYGGGKIIFIDTENTRPRDR-----LRDIADRFNVDR-----VLDNVL 189
QY 144 YFRICSTEQIAVINYM-EKFLREHKDRIYIIVDFHFRDFE--DLAIFRIVSGSL 199
DB 190 YARAYTSEHQMELLDYVAAKHEHAGIFKLLIIDSIMALFRVDSGKGLAEKQKLAOM 249
QY 200 SLKIMIAKTYNLAIVLLNQVTFKTEGSGFQLTLAL-----GDSWSHSCNRLILH-W 251
DB 250 LSLRLQKISEYVNVAVFVNTQMT---ADPGATMTFOADPKPKPIGHIILAHSTRTISLR 306
QY 252 NGNERVAHLDKPSPLPVAAPYAVTGGKIRDA 283
DB 307 RGLRLAKIYDSPENPENEAFTTAGIGIDA 338

```

```

RESULT 9
DMCL_LILLO STANDARD; PRT; 349 AA.
AC P37384;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)

```





OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96363911; PubMed=8725225;  
 RA Diemer A.C., Fink G.R.;  
 RT "DLH1 is a functional Candida albicans homologue of the  
 RL meiosis-specific gene DMCL1";  
 RL Genetics 143:769-776(1996).  
 CC -1- FUNCTION: REQUIRED FOR MEIOTIC RECOMBINATION, SYNAPTONEMAL COMPLEX  
 CC FORMATION AND CELL CYCLE PROGRESSION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO  
 CC PROKARYOTIC RECA PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U39808; AAC49400.1; -  
 DR PROSITE: PS50162; RECA\_2; 1.  
 DR PROSITE: PS50163; RECA\_3; 1.  
 KW Meiosis; Cell cycle; ATP-binding; Nuclear protein.  
 FT NP\_BIND 112 119 ATP (POTENTIAL).  
 FT SEQUENCE 324 AA; 35281 MW; 95CD769514D9CF82 CRC64;  
 SQ  
 Query Match 19.9%; Score 304; DB 1; Length 324;  
 Best Local Similarity 33.2%; Pred. No. 5.5e-19;  
 Matches 87; Conservative 38; Mismatches 99; Indels 38; Gaps 7;  
 OY 34 HTTSGDLDNDILGGGICKEVEYTGVPVGKTOGLQIALNVOIPVECGLGKAVYI 93  
 Db 86 HTTSGKQDEILGGGISTEYGEFRCKTQCHTLCAADLPDMGSGEGKAVYI 145  
 OY 94 DTGSGFWEVRYQVIAE--GCIKDILHFRPHSHKSSSVQKQLOPERLADIIYFRICSYT 151  
 Db 146 DTGTFPRPRIRSIARVGDADIC-----LENISYARALNSE 183  
 OY 152 EQAVINVMKFLREHKDRIYIDSVTFHRODFE--DLALRTVLSGLSLKMKIAK 208  
 Db 184 HQLELEQGLNELAE-GTFRLLIVDSIMACFRVDSGKELNROOKNQHLSNLTVAE 242  
 OY 209 TYNLAVVLNQVTTKFTGSGFQTLAL-----GDSMHSCTNRLILH-WNGNERYAHL 260  
 Db 243 DYNIAVFLNQVOS--DPGASALFAAADGRKRVGGHVLHAASNTRILLKRGGERVAKL 300  
 OY 261 DKSPSLPVASAPYAVTGKGI RD 282  
 Db 301 ODSPNMEKECVYVIGEGIKD 322  
 RESULT 14  
 ID RAS1\_USTMA STANDARD; PRT; 339 AA.  
 AC 099133;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE DNA REPAIR PROTEIN RAD51.  
 GN RAD51.  
 OS Ustilago maydis (Smut fungus).  
 OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes; Ustilago.  
 CC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.  
 OX NCBI\_TaxID=5270;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97225207; PubMed=9071580;  
 RA Ferguson D.O., Rice M.C., Rendi M.H., Kotani H., Kmiec E.B.,  
 RA Holloman W.K.;

RT "Interaction between Ustilago maydis REC2 and RAD51 genes in DNA  
 RT repair and mitotic recombination.";  
 RL Genetics 145:243-251(1997).  
 CC -1- FUNCTION: REQUIRED BOTH FOR RECOMBINATION AND FOR THE REPAIR OF  
 CC DNA DAMAGE CAUSED BY X-RAYS (BY SIMILARITY).  
 CC -1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO  
 CC PROKARYOTIC RECA PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U62484; AAC61878.1; -  
 DR InterPro: IPR000445; -  
 DR Pfam: PF00633; HHH; 1.  
 DR PROSITE: PS50162; RECA\_2; 1.  
 DR PROSITE: PS50163; RECA\_3; 1.  
 KW DNA damage; DNA repair; ATP-binding; DNA recombination.  
 FT NP\_BIND 127 134 ATP (POTENTIAL).  
 FT SEQUENCE 339 AA; 36396 MW; DOE3676272FB0254 CRC64;  
 SQ  
 Query Match 19.6%; Score 299.5; DB 1; Length 339;  
 Best Local Similarity 32.4%; Pred. No. 1.4e-18;  
 Matches 84; Conservative 46; Mismatches 98; Indels 31; Gaps 6;  
 OY 35 ITTSGDLDNDILGGGICKEVEYTGVPVGKTOGLQIALNVOIPVECGLGKAVYI 94  
 Db 102 ITTSGKNDILGGGIMETGISTEYGEFRCKTQCHTLAVTCOLPDMGSGEGKCLYID 161  
 OY 95 TEGSGFWEVRYQVIAE--GCIKDILHFRPHSHKSSSVQKQLOPERLADIIYFRICSYT 154  
 Db 162 TETTFPRPRIRSIARVGDADIC-----LENISYARALNSE 201  
 OY 155 AVINVMKFLREHKDRIYIDSVTFHRODFE--DLALRTVLSGLSLKMKIAKTYN 211  
 Db 202 QLMQASAMMAESR-FSLIYDLSLSYRTDFSGRGLSARQMLAFLGLRLRDLDFG 260  
 OY 212 LAVVLNQVTTKFTGSGFQTLAL-----GDSMHSCTNRLILH-WNGNERYAHLKSPS 265  
 Db 261 VAVVITNQVAAQ--VDGATAFADAKKPIGNTIVAHASTRLSLRKGNGRIGRIADSPC 319  
 OY 266 LPVASAPYAVTGKGI RD 284  
 Db 320 LPEDAVFAISPEGITDPV 338  
 RESULT 15  
 ID DMCL1\_SCHPO STANDARD; PRT; 332 AA.  
 AC 042634; 042880;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE MEIOTIC RECOMBINATION PROTEIN DMCL1.  
 GN DMCL1 OR DMP1 OR SPAC8E11.03C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC Schizosaccharomycetaceae; Schizosaccharomycetaceae;  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND FUNCTION.  
 RX MEDLINE=20368631; PubMed=10908327;  
 RA Fukushima K., Tanaka Y., Nabeshima K., Yoneki T., Tougan T.,  
 RA Tanaka S., Nojima H.;  
 RT "Mcl1 of Schizosaccharomycetes pombe plays a role in meiotic  
 RT recombination.";  
 RL Nucleic Acids Res. 28:2709-2716(2000).





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2001, 16:27:27 ; Search time 15.99 Seconds  
(without alignments)  
1400.583 Million cell updates/sec

Title: US-09-537-654-2

Perfect score: 1530  
Sequence: 1 MDQSGSRNGPQOKTVSGAQ.....VTGKGRDAVSSNKRRAVT 294

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR.68:\*\*  
1: PIR1:\*\*  
2: PIR2:\*\*  
3: PIR3:\*\*  
4: PIR4:\*\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	988	64.6	332	2	probable RAD51C-11
2	417	27.3	319	2	rad51 protein - Sul
3	387.5	25.3	324	2	DNA repair protein
4	354.5	23.0	317	2	DNA repair protein
5	352.5	23.0	311	2	DNA repair protein
6	345	22.5	356	2	meiosis-specific r
7	337	22.0	334	2	meiosis-specific r
8	314.5	20.6	340	2	UC4191
9	313	20.5	340	2	meiosis-specific r
10	313	20.5	349	2	hypothetical 38.3K
11	312	20.4	345	2	RecA/Rad51/DMC1-11
12	312	20.4	352	2	DNA repair protein
13	309	20.2	357	2	rad51 protein, sho
14	309	20.2	391	2	hypothetical prote
15	304	19.9	324	2	DMC1/LIM15 homolog
16	298.5	19.5	336	2	meiotic recombina
17	297.5	19.4	332	2	meiotic recombina
18	293.5	19.2	386	2	DNA repair protein
19	290	19.0	342	2	DMC1/LIM15 homolog
20	287.5	18.8	339	2	DNA repair protein
21	287.5	18.8	343	2	rad51 protein - Hal
22	281.5	18.4	339	2	rad51 protein homo
23	280.5	18.3	339	2	recombination/rep
24	277.5	18.1	400	2	rad51 protein - ye
25	276	18.0	365	2	rad51 protein homo
26	272.5	17.8	342	2	rad51 protein homo
27	265.5	17.4	353	2	rad51 homolog mei-
28	260.5	17.0	354	2	DNA repair protein

## ALIGNMENTS

## RESULT 1

F84888 probable RAD51C-like DNA repair protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: F84888

R/lin, X.J. Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.I., Town, C.D., Fujii, C.Y.

esus, D., J. Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M., Venter

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: F84888

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-332 <STO>

A:Cross-references: GB:AE002093; NID:g2583126; PIDN:AAB82635.1; GSPDB:GNO0139

C:Genetics:

A:Gene: At2g45280

A:Map position: 2

Query Match 64.6%; Score 988; DB 2; Length 332;

Best Local Similarity 66.5%; Pred. No. 2.9e-77;

Matches 183; Conservative 51; Mismatches 39; Indels 2; Gaps 2;

QY 19 AONAMDFSDLSOKATTTGSGDLNDLGGCHCKEYETIGVPGVGTQGLAINVO 78

Db 59 AKNAMDMLHEESLPRTTSCSDNLTGGISCRDTGTVGPGVGTQGLAINVO 118

QY 79 IPVCGGLGKAVYIDPESFMYERVYQIAEGCIRDLIEHPSPHSHEKSSVQKQLOPERF 138

Db 119 IPRCGGLGKAVYIDPESFMYERVYQIAEGCIRDLIEHPSPHSHEKSSVQKQLOPERF 178

QY 139 LADYFRICSYEQIAVINMEKREKRDVRIYIDSYTFHRQDEFLAIRYLSG 198

Db 179 LENIFRVCSTYEQIALVNHLEKFEISEKNDV-VVIDYSTTFHRQDYDLAORTYLSG 237

QY 199 LSLKMLAKTYNLAVALVNOVTKFESEFQTLALGDSMSHSCYRLLIHNNGNRYA 258

Db 238 MALFEMKLAKFSLAVLVNOVTKFESEFQTLALGDSMSHSCYRLLIHNNGNRYA 297

QY 259 HLDKSPSLPVASAPYAVTGKGRDAVSSNKRRAVT 293

Db 298 YIDKSPSLPVASAPYAVTGKGRDAVSSNKRRAVT 331

RESULT 2

D72766 probable rada protein APE0119 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000







```

Query Match          20.5%; Score 312; DB 2; Length 349;
Best Local Similarity 33.3%; Pred. No.2.9e-19;
Matches 92; Conservative 46; Mismatches 104; Indels 32; Gaps 7

OY 15 YVSGAQNAMDMSDELISKHITTTGSDLLDILGGHCKEVEITIGVPGVGKTOLGIOLA 74
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::
DB 97 YITGS----DVLKRRSVIRITTTGSDALDELGGVETSAITAEAFGRSGKTOLAHITLC 152
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::

OY 75 INVQIPVECGGIGGAGNAVYIDTEGSEFMVERVYIOIAGCIRDLIEHPPHSEKSSSVOKOLO 134
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::
DB 153 VSTOLPVMHGNGKGVAYIDTEGTFPRDRIIVEIAERFGMD-----ASAV----- 196
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::

OY 135 PERFLADIFYFCSTGTEOIAVINMEKFLREHKDVRIVTIDSYTFHFRODE---DLAL 191
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::
DB 189 PGAVLDNIITTYARATYIEHOYNLLGLAKMSE--EPFRLLIDVSVALRPVDSGREGELAD 247
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::

OY 192 RTRVLISGLSLKMLKIATKYNLNAVLLNOVTTKFTGGSF---OLTIALGDSWSHSCSTNRLI 248
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::
DB 248 RQOKLAQMLSLRLTIKIAEENFNAVYMTNQVIADPGGGMFLSDPKKPAAGGHVLAHAATVRLM 311
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::

OY 249 LH-WNGENRYAHLDKSPSLPVASAPYAVTGGCIRDA 283
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::
DB 312 LKKGGEQRVCKIFDAPNLPESAFAVFOITPAGVADA 347
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::

RESULT 11
T08838
Reca/Rad51/DMC1-like protein - soybean
C:Species: Glycine max (soybean)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 23-Jul-1999
C:Accession: T088838
R:Rad1, M.2.; Finer, J.J.
submitted to the EMBL Data Library, August 1996
A:Description: RecA like gene from soybean.
A:Reference number: Z16485
A:Accession: T08838
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-345 <HAD>
A:Cross-references: EMBL:U66836; NID:g1518156; PID:g1518157
C:Function:
A:Description: recombinase
C:Superfamily: yeast DNA repair protein RAD51

Query Match          20.4%; Score 312; DB 2; Length 345;
Best Local Similarity 32.6%; Pred. No.3.4e-19;
Matches 90; Conservative 46; Mismatches 108; Indels 32; Gaps 6;

OY 15 YVSGAQNAMDMSDELISKHITTTGSDLLDILGGHCKEVEITIGVPGVGKTOLGIOLA 74
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::
DB 93 YITGS----DALTKRKSIVIRITTTGSDALDELGGVETSAITAEAFGRSGKTOLAHITLC 148
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::

OY 75 INVQIPVECGGIGGAGNAVYIDTEGSEFMVERVYIOIAGCIRDLIEHPPHSEKSSSVOKOLO 134
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::
DB 149 VSTOLPVMHGNGKGVAYIDTEGTFPRDRIIVEIAE-----RFGMD 188
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::

OY 135 PERFLADIFYFCSTGTEOIAVINMEKFLREHKDVRIVTIDSYTFHFRODE---DLAL 191
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::
DB 189 PGAVLDNIITTYARATYIEHOYNLLGLAKMSE--EPFRLLIDVSVALRPVDSGREGELAD 247
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::

OY 192 RTRVLISGLSLKMLKIATKYNLNAVLLNOVTTKFTGGSF---OLTIALGDSWSHSCSTNRLI 248
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::
DB 248 RQOKLAQMLSLRLTIKIAEENFNAVYMTNQVIADPGGGMFLSDPKKPAAGGHVLAHAATVRLM 307
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::

OY 249 LH-WNGENRYAHLDKSPSLPVASAPYAVTGGCIRDA 283
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::
DB 308 FRKKGGEQRICKVFDAPNLPESAFAVFOITAGIADA 343
   |::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::

```

C:Species: Methanococcus jannaschii  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
C:Accession: E64408; S71095; S71096  
R:Butl, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak  
R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,  
Iron, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese  
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc  
A:Reference number: A64300; MUID:96337999  
A:Accession: E64408  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-352 <Bul>  
A:Cross-references: GB:067531; GB:L77117; NID:q2826342; PIDD:AA98875.1; PID:915915533  
R:Sandler, S.J.; Satlin, L.H.; Samra, H.S.; Clark, A.J.  
Nucleic Acids Res. 24, 2125-2132, 1996  
A:Title: recA-like genes from three archaean species with putative protein products s  
A:Reference number: S71093; MUID:96251095  
A:Accession: S71095  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 33-352 <SA2>  
A:Cross-references: EMBL:U45312  
C:Genetics:  
A:Gene: rada  
A:Map position: FOR791863-792921  
C:Superfamily: yeast DNA repair protein RAD51

Query Match similarity 20.4%; Score 312; DB 2; Length 352;  
Best Local similarity 31.0%; Pred. 3.5e-19;  
Matches 92; Conservative 47; Mismatches 106; Indels 52; Gaps 8;

```
OY      1 MGDGSGSRNGPQOKYKVSQAONAMDMFSDLSOKHITGSGDLNLLGCGICKEVTEIYG 60
      1:111:      : : : : : : : : : : : : : : : : : : : : : : : : :
Db      93 LGFKSGT-----EVLSSQRKNIM-----KLTSGSKNLDLITLGGLESQVTEFAG 136
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      61 VPGVKKTQLOGLQALINQVLPYECGSLG-----KAVYIDEGSMVRRVYQIAGC 111
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      137 MFGSKGTQIADQACVNIQCPERIVADDAIKDEILNEPAAVYIDTEGTFPRPRIVQMAEAL 196
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      112 IRDLIEHPHSHEKSSVOKOLOPERFLADLYPRICSYTQIQAIVINMEKFLREHDDR 171
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      197 -----GLDGEVLENNIFVARAYISDMQMLAEVENVLIIDGNRIK 236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      172 IVIDSVTEHFRQDF--EDLALRTRVLGSLKMLKTIKTYNLAVALLNQVTK--FTE 226
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      237 LVIVDSLTSTFRTETXIGRGKLAERQOKIGRHMATLNKLADLYINCVCVIVTQVAPDALF 296
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      227 GSFQTLTALGSMWSHSCNRLILH-WGNENERVAHLDSPSLPVASAPYAVYQKRIKD 282
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      297 GPSEQAIC-GHIVGHAATFRIFLKRAGKDKRVAKLYDSPHLDPDAEMFRTEKGIHD 352
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 13

T37305

rad51 protein, short isoform - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jul-2000

C:Accession: T37305; T43059

R:Takanami, T.; Sato, S.; Ishihara, T.; Katsura, I.; Takahashi, H.; Higashitani, A.

DNA Res. 5, 373-377, 1998

A:Title: Characterization of a *Caenorhabditis elegans* recA-like gene Ce-rdh-1 involve

A:Reference number: Z21672; MUID:99156232

A:Accession: T37305

A:Status: preliminary; translated from GB/EMBL/DBAJ

A:Molecule type: mRNA

A:Residues: 1-357 <TRAK>

A:Cross-references: EMBL:AB011382; NID:g2913896; PIDN:BAA24982.1; PID:g2913897

R:Rinaldo, C.; Ederle, S.; Rocco, V.; La Volpe, A.

Mol. Gen. Genet. 260, 289-294, 1998

Title: The *Caenorhabditis elegans* Rad51 homolog is transcribed into two alternative

A:Reference number: 222307; MDID:99077299  
A:Accession: T43059  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-357 <RIN>  
A:Cross-references: EMBL:AF061201; NID:g3786401; PID:g3786402  
C:Genetics:  
C:Gene: rad51  
C:Superfamily: yeast DNA repair protein RAD51

Query Match	20.28%	Score 309;	DB 2;	Length 357;
Best Local Similarity	33.78%	Pred. No. 6.5e-19;		
Matches	87;	Conservative	36;	Mismatches 105; Indels 30; Gaps 6

```

Oy 35 ITTGGGDLNDLIDGGGHHCKEYTEIGVPGVKOTLGJLAINVOIPECCGGLGKAVYID 94
Db 118 IRTGSASDLRLLGGIENGSIIEYGEYRTKQTOLCSHLAVLCQPLPDMGGGGEKCYAID 177
Oy 95 TEGSMVVRVYQIAEGCTRIODLLEHFRPHSHKSSSVQKQLOQPERFLADIIYFRICSTYEQI 154
Db 178 TNATFRPRRIIAIA-----RNMSSAHVLEIVAVARAENSEHM 217
Oy 155 AVINMEKFELEBKHDRIVIDSYVTFHERODE--DLALFTRVLSISLKIIMAKTYN 211
Db 218 ALIIAGAGMMSESR-YAVYIVDCATAHFRNNRYITRGDLAEQMKLSAFLCMIKLADIEYG 276
Oy 212 LAVVLLNOVTRKTEGGS--FOLLTAL--GDSWGHSTCNRLILH-WNGNERIAHLDSPS 265
Db 277 VAVIITNOVVAOVGSGAMFOADAKKPIGSHIIAHMSTRIYLRKKGGENNVAKMOSPN 336
Oy 266 LPVASAPYAAVNGKIRDA 283
Db 337 LPFAEATYSITNMHGIEDA 354

```

RESULT 14  
T26822  
Hypothetical protein Y43C5A.6 - *Caenorhabditis elegans*  
C.Species: *Caenorhabditis elegans*  
C.Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
C.Accession: T26822  
R.White, S.  
submitted to the EMBL Data Library, June 1998  
A.Reference number: Z2072  
A.Accession: T26822  
A.Status: preliminary; translated from GB/EMBL/DDAJ  
A.Molecule type: DNA  
A.Residues: 1-391 <MIL>  
A.Cross-references: EMBL:AL023638; PIDN:CAAI9500.1; GSPDB:GN00022; CESP:Y43C5A.6  
A.Experimental source: clone Y43C5A  
A.Genetics:  
A.Gene: CESP:Y43C5A.6  
A.Map position: 4  
A.Introns: 31/3; 125/3; 174/3; 227/2; 258/1; 350/2  
C:Superfamily: yeast DNA repair protein RAD51

RESULT 15  
S70390  
DMC1/LIM15 homolog 1 - Yeast (*Candida albicans*)  
C.Species: *Candida albicans*  
C.Date: 06-Dec-1996 #sequence\_revision 13-Mar-1997 #text\_change 29-Sep-1999  
C.Accession: S70390  
R.Dienert, A.C.; Fink, G.R,  
Genetics 143, 769-776, 1996  
A.Title: Dhl1 is a functional *Candida albicans* homologue of the meiosis-specific genea  
A.Reference number: S70390; MUID:96363911  
A.Accession: S70390  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-324 <DIE>  
A.Cross-References: EMBL:U39808; NID:G1145715; PID:G1145716  
C.Genetics:  
A:Gene: Dhl1  
A:Introns: 35/3; 67/3  
C:Superfamily: yeast DNA repair protein RAD51

Query Match	19.98;	Score 304;	DB 2;	Length 324;
Best Local Similarity	33.2%;	Pred. No. 1.5e-18;		
Matches	87;	Conservative	38;	Mismatches 99; Indels 38; Gaps 7

```
Search completed: October 1, 2001, 16:28:37
Job time: 70 sec
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 05:00:54 ; Search time 124.96 Seconds  
(without alignments)  
7406.586 Million cell updates/sec

Title: US-09-537-654-1

Perfect score: 1474  
Sequence: 1 tgcaccacagctccgcact.....aaaaaaaaaaaaaaaa 1474

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_0601:\*

- 1: /SIDSL1/gcgdata/geneseq/geneseqn/NA1980.DAT:\*
- 2: /SIDSL1/gcgdata/geneseq/geneseqn/NA1981.DAT:\*
- 3: /SIDSL1/gcgdata/geneseq/geneseqn/NA1982.DAT:\*
- 4: /SIDSL1/gcgdata/geneseq/geneseqn/NA1983.DAT:\*
- 5: /SIDSL1/gcgdata/geneseq/geneseqn/NA1984.DAT:\*
- 6: /SIDSL1/gcgdata/geneseq/geneseqn/NA1985.DAT:\*
- 7: /SIDSL1/gcgdata/geneseq/geneseqn/NA1986.DAT:\*
- 8: /SIDSL1/gcgdata/geneseq/geneseqn/NA1987.DAT:\*
- 9: /SIDSL1/gcgdata/geneseq/geneseqn/NA1988.DAT:\*
- 10: /SIDSL1/gcgdata/geneseq/geneseqn/NA1989.DAT:\*
- 11: /SIDSL1/gcgdata/geneseq/geneseqn/NA1990.DAT:\*
- 12: /SIDSL1/gcgdata/geneseq/geneseqn/NA1991.DAT:\*
- 13: /SIDSL1/gcgdata/geneseq/geneseqn/NA1992.DAT:\*
- 14: /SIDSL1/gcgdata/geneseq/geneseqn/NA1993.DAT:\*
- 15: /SIDSL1/gcgdata/geneseq/geneseqn/NA1994.DAT:\*
- 16: /SIDSL1/gcgdata/geneseq/geneseqn/NA1995.DAT:\*
- 17: /SIDSL1/gcgdata/geneseq/geneseqn/NA1996.DAT:\*
- 18: /SIDSL1/gcgdata/geneseq/geneseqn/NA1997.DAT:\*
- 19: /SIDSL1/gcgdata/geneseq/geneseqn/NA1998.DAT:\*
- 20: /SIDSL1/gcgdata/geneseq/geneseqn/NA1999.DAT:\*
- 21: /SIDSL1/gcgdata/geneseq/geneseqn/NA2000.DAT:\*
- 22: /SIDSL1/gcgdata/geneseq/geneseqn/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1474	100.0	1474	22	MAIZE RAD51 orthol
2	1305.4	88.6	1333	22	MAIZE RAD51 orthol
3	1243.2	84.3	1459	22	MAIZE RAD51 orthol
4	178.4	12.1	1325	21	HTRM clone 2398682
5	90.2	6.1	936	22	OLIGONUCLEOTIDE D1
6	90.2	6.1	936	22	OLIGONUCLEOTIDE D1
7	90.2	6.1	936	22	OLIGONUCLEOTIDE D1
8	90.2	6.1	936	22	OLIGONUCLEOTIDE D2
9	90.2	6.1	936	22	OLIGONUCLEOTIDE D2
10	89.8	6.1	936	22	OLIGONUCLEOTIDE D1
11	89.8	6.1	936	22	OLIGONUCLEOTIDE D1

12	89.8	6.1	936	22	AAF58254	OLIGONUCLEOTIDE D1
13	89.8	6.1	936	22	AAF58257	OLIGONUCLEOTIDE D1
14	89.8	6.1	936	22	AAF58259	OLIGONUCLEOTIDE D2
15	89.8	6.1	936	22	AAF58262	OLIGONUCLEOTIDE D2
16	89.8	6.1	938	22	AAF58255	OLIGONUCLEOTIDE D1
17	65.8	4.5	1020	20	AAZ11866	ZMRAD51A nucleotid
18	65.8	4.5	1568	20	AAZ11868	ZMRAD51A cDNA clon
19	65	4.4	1585	22	AAZ02318	ZMRAD51B nucleotid
20	64.4	4.4	1020	20	AAZ11870	ZMRAD51B cDNA clon
21	64.4	4.4	1574	20	AAZ11872	ZMRAD51B cDNA clon
22	62.8	4.3	1219	22	AAZ02316	Glycine max RAD51
23	57.8	3.9	1682	16	AAQ91763	Human Rad51 gene,
24	57.8	3.9	2229	21	AAA37710	Human Rad51 mRNA.
25	55.8	3.8	1362	21	AAZ47934	Arabidopsis thalia
26	55.8	3.8	1408	16	AAQ91761	Mouse Rad51 gene,
27	55.8	3.8	1755	15	AAQ64088	Mouse gene particl
28	53.8	3.6	3776	17	AAZ42903	TRP-1 protein codi
29	53.8	3.6	3994	21	AAZ76475	Human OREF2030
30	52.8	3.6	2639	21	AAZ58108	Human PRO1788 nucl
31	52.8	3.6	2639	21	AAA37137	Human PRO1788 (UNQ
32	52.8	3.6	2639	22	AAF54480	Probe #56 used in
33	52.4	3.6	2422	21	AAZ65112	Human OREF ORF1667
34	52.2	3.5	887	21	AAZ78089	Human cancer assoc
35	52	3.5	2737	21	AAZ64478	Human H38087 (clon
36	51	3.5	3447	21	AAZ76035	Human OREF ORF1590
37	50.8	3.4	876	21	AAZ59201	Human secreted pro
38	49.8	3.4	1522	21	AAZ93480	Human secreted pro
39	49.8	3.4	1641	17	AAZ15758	Human interleukin-
40	49.8	3.4	4091	11	AAZ004035	Stem cell leukemia
41	49.8	3.4	4199	11	AAZ05330	Stem cell leukemia
42	49.2	3.3	1483	21	AAZ65060	Membrane-bound pro
43	49.2	3.3	1484	22	AAZ44206	Human PRO1105 (UNQ
44	48.6	3.3	443	21	AAA75749	cDNA encoding a de
45	48.6	3.3	553	21	AAZ79990	Human secreted pro

#### ALIGNMENTS

RESULT	1	
AD02118	standard; cDNA; 1474 BP.	
ID	AD02118	
AC	AD02118:	
XX		
XX	28-MAR-2001 (first entry)	
XX		
XX	Maize RAD51 orthologue #1 cDNA.	
DE		
XX	Maize: RAD51 orthologue; RAD52 epistasis group; double strand break; DSB;	
KW	homologous recombination; transformation; gene targeting;	
KW	transgenic plant; DNA repair; ss.	
OS		
XX	Zea mays.	
XX		
PH	Key	Location/Qualifiers
FT	CDS	310..1194
FT		/*tag= a
FT		/product= "Maize RAD51 orthologue protein #1"
XX		
PN	W0200068370-A2.	
XX		
PD	16-NOV-2000.	
XX		
PF	05-APR-2000: 2000MO-US09010.	
XX		
PR	05-MAY-1999: 99US-0132582.	
XX		
PA	(PION-) PIONEER HI-BRED INT INC.	
XX		
PI	Mahajan PB, Shi J;	
XX		
DR	WPI; 2001-016092/02.	

DR P-PSDB; AAY1974.

XX New maize RAD51 polynucleotides, useful for e.g. controlling  
PT recombination or transformation efficiency in plants, or as probes or  
PT amplification primers for detecting, quantifying or isolating gene  
PT transcripts

XX Claim 1, Page 68-69; 76pp; English.

XX The present sequence is a maize RAD51 orthologue #1 cDNA. RAD51 is one  
CC of the genes of the RAD52 epistasis group that is involved in double  
CC strand breaks (DSBs) repair by homologous recombination. Control of  
CC homologous recombination by modulating RAD51 is used for improving  
CC transformation and gene targeting in transgenic plants. The RAD51  
CC nucleotide may be used as probes or amplification primers for detecting,  
CC quantifying or isolating gene transcripts, in detecting deficiencies in  
CC the mRNA level during screening for desired transgenic plants, for  
CC detecting gene mutations, for monitoring upregulation of expression or  
CC changes in enzyme activity, for detecting any number of allelic variants,  
CC orthologues or paralogues of the gene, or for site directed mutagenesis  
CC in eukaryotic cells. It may also be used for recombinant expression of  
CC its encoded polypeptide, or for use as immunogen in preparing and/or  
CC screening of antibodies, and in sense or antisense suppression of one or  
CC more genes in a host cell, tissue or plant. The RAD51 proteins may be  
CC used in assays to agonise or antagonise the enzyme function, or as  
CC immunogens or antigens to obtain antibodies.

XX Sequence 1474 BP; 415 A; 337 C; 369 G; 353 T; 0 other;

Query Match 100.0%; Score 1474; DB 22; Length 1474;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgaccacgctccgctcacttacttccacatgctgagcttgcgttgc 60  
DB 1 tcgaccacgctccgctcacttacttccacatgctgagcttgcgttgc 60  
QY 61 cggagcccaaaagcgcggtgagccgagccgagacgagcgcgagctccc 120  
DB 61 cggagcccaaaagcgcggtgagccgagccgagacgagcgcgagctccc 120  
QY 121 cctaaagcgacgagcggtgagccgagccgagacgagcgcgagctccc 180  
DB 121 cctaaagcgacgagcggtgagccgagccgagacgagcgcgagctccc 180  
QY 181 gagcggttggagagatgagagtgagagtgagagtgagagtgagagtgag 240  
DB 181 gagcggttggagagatgagagtgagagtgagagtgagagtgagagtgag 240  
QY 241 gatattgaaatccgacatgcttctctctcccaatcagtaaccatctc 300  
DB 241 gatattgaaatccgacatgcttctctctcccaatcagtaaccatctc 300  
QY 301 gcaatcacatggagatcaatctgctctagaatgagacacagaagttag 360  
DB 301 gcaatcacatggagatcaatctgctctagaatgagacacagaagttag 360  
QY 361 ggaagccgaatgctgagatgcttctctgaatgagctgtcagagaaacatc 420  
DB 361 ggaagccgaatgctgagatgcttctctgaatgagctgtcagagaaacatc 420  
QY 421 ggttctggtgacatcaatgacatctggtgagagtgacatgacaaagttag 480  
DB 421 ggttctggtgacatcaatgacatctggtgagagtgacatgacaaagttag 480  
QY 481 atcgtgtgctccaggggtgttaaactcaactgagggatcaactgacataatga 540  
DB 481 atcgtgtgctccaggggtgttaaactcaactgagggatcaactgacataatga 540  
QY 541 caaatcccaagtgagtgctgcttggtgagaaagcagttatatatagacagaagc 600  
DB 541 caaatcccaagtgagtgctgcttggtgagaaagcagttatatatagacagaagc 600

QY 601 agtttcagtgtagacgctgcttaccagatgctgaaggggtattagggacatctggag 660  
DB 601 agtttcagtgtagacgctgcttaccagatgctgaaggggtattagggacatctggag 660  
QY 661 cacttccgacagacatgagaggttctctcttctcccaaaacatcacagctgagcgt 720  
DB 661 cacttccgacagacatgagaggttctctcttctcccaaaacatcacagctgagcgt 720  
QY 721 ttctctgaggtatctatcttacttccgagatgacatgacacgaatctgacatgata 780  
DB 721 ttctctgaggtatctatcttacttccgagatgacatgacacgaatctgacatgata 780  
QY 781 aactacaatggagaggttccctcagagacataaagatgtgctatgattatgagtagt 840  
DB 781 aactacaatggagaggttccctcagagacataaagatgtgctatgattatgagtagt 840  
QY 841 gtaacttccacttccgacaaagatttgaagatctgacactgagagacagagtgtaagt 900  
DB 841 gtaacttccacttccgacaaagatttgaagatctgacactgagagacagagtgtaagt 900  
QY 901 ggtattatcattgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 960  
DB 901 ggtattatcattgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 960  
QY 961 aaccaagtcactactaaattacagaaagggtcatttcaatgactctgtctaggtgac 1020  
DB 961 aaccaagtcactactaaattacagaaagggtcatttcaatgactctgtctaggtgac 1020  
QY 1021 agctgttccactactacacagacggttgaatctgacactgagagtggaagacagatc 1080  
DB 1021 agctgttccactactacacagacggttgaatctgacactgagagtggaagacagatc 1080  
QY 1081 gcaatcttgaatgaatgcttccctcacttccagtagcctcagcagcagatgagtagagc 1140  
DB 1081 gcaatcttgaatgaatgcttccctcacttccagtagcctcagcagcagatgagtagagc 1140  
QY 1141 aaagggatgagagtgctgtgaggtcacaacacacagagcgagccgagtaagtagcatct 1200  
DB 1141 aaagggatgagagtgctgtgaggtcacaacacacagagcgagccgagtaagtagcatct 1200  
QY 1201 tgggtgtaagcactgtatgttccactacgctctctgacgcttcttccgacatgagcttct 1260  
DB 1201 tgggtgtaagcactgtatgttccactacgctctctgacgcttcttccgacatgagcttct 1260  
QY 1261 ggaactagtgagtgagagtgagagatgacacatttgttgaattcctcagttgctgtgc 1320  
DB 1261 ggaactagtgagtgagagtgagagatgacacatttgttgaattcctcagttgctgtgc 1320  
QY 1321 cgttggcttaccacaaacttaagagagagtaataacacagacagcgttaataagtg 1380  
DB 1321 cgttggcttaccacaaacttaagagagagtaataacacagacagcgttaataagtg 1380  
QY 1381 ttgtgttctggaacatctggtccatgctacatcagtaagagcctataatagcgggcaaaa 1440  
DB 1381 ttgtgttctggaacatctggtccatgctacatcagtaagagcctataatagcgggcaaaa 1440  
QY 1441 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1474  
DB 1441 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1474

RESULT 2

AAD02120 standard; cDNA; 1333 BP.

XX AAD02120;

XX 28-MAR-2001 (first entry)

XX Maize RAD51 orthologue #3 cDNA.

KW Maize; RAD51 orthologue; RAD52 epistasis group; double strand break; DSB;

KW homologous recombination; transformation; gene targeting;  
transgenic plant; DNA repair; ss.

XX Zea mays.

XX Key Location/Qualifiers  
XX FT CDS 169..1053  
XX FT /\*tag= a  
XX FT /product= "Maize RAD51 orthologue protein #3"

XX W0200068370-A2.

XX 16-NOV-2000.

XX 05-APR-2000; 2000MO-US09010.

XX 05-MAY-1999; 99US-0132582.

XX (PION-) PIONEER HI-BRED INT. INC.

XX Mahajan PB, Shi J.

XX WPI; 2001-016092/02.

XX P-PSDB; AAY71976.

PT New maize RAD51 polynucleotides, useful for e.g. controlling  
PT recombination or transformation efficiency in plants, or as probes or  
PT amplification primers for detecting, quantifying or isolating gene  
PT transcripts.

XX Claim 1; Page 72-74; 76pp; English.

CC The present sequence is a maize RAD51 orthologue #3 cDNA. RAD51 is one  
CC of the genes of the RAD52 epistasis group that is involved in double  
CC strand breaks (DSBs) repair by homologous recombination. Control of  
CC homologous recombination by modulating RAD51 is used for improving  
CC transformation and gene targeting in transgenic plants. The RAD51  
CC nucleotide may be used as probes or amplification primers for detecting,  
CC quantifying or isolating gene transcripts, in detecting deficiencies in  
CC the mRNA level during screening for desired transgenic plants, for  
CC detecting gene mutations, for monitoring upregulation of expression or  
CC changes in enzyme activity, for detecting any number of allelic variants,  
CC orthologues or paralogues of the gene, or for site directed mutagenesis  
CC in eukaryotic cells. It may also be used for recombinant expression of  
CC its encoded polypeptide, or for use as immunogen in preparing and/or  
CC screening of antibodies, and in sense or antisense suppression of one or  
CC more genes in a host cell, tissue or plant. The RAD51 proteins may be  
CC used in assays to agonise or antagonise the enzyme function, or as  
CC immunogens or antigens to obtain antibodies.

XX Sequence 1333 BP; 377 A; 284 C; 326 G; 345 T; 1 other;

Query Match 88.6%; Score 1305.4; DB 22; Length 1333;

Best Local Similarity 98.6%; Pred. No. 0;  
Matches 1315; Conservative 1; Mismatches 17; Indels 0; Gaps 0;

QY 142 cgaagtaagcgtgctggtgcgcacacgaagagcgtggtgagagcagata 201  
DB 1 cgaagtaagcgtgctggtgcgcacacgaagagcgtggtgagagcagata 60  
QY 202 tgaagagtgagagtggtacacagcgtgctgagatgataactaaatccgcatgca 261  
DB 61 tgaagagtgagagtggtacacagcgtgctgagatgataactaaatccgcatgca 120  
QY 262 gttctctcttcccccaatacgaacacgtcctcaagtgtgcaatcacatggagatcaa 321  
DB 121 gttctctcttcccccaatacgaacacgtcctcaagtgtgcaatcacatggagatcaa 180  
QY 322 tctgctctagaatgagcacaacagaatgactttcagagccagaaatgctgctgagat 381  
DB 181 tctgctctagaatgagcacaacagaatgactttcagagccagaaatgctgctgagat 240

QY 382 atgtctctgtagtgcgtgtcacagaaacacatcactactgtgtctgtgacatcaatgac 441  
DB 241 atgtctctgtagtgcgtgtcacagaaacacatcactactgtgtctgtgacatcaatgac 300  
QY 442 atactgtgtgctgagatcactgtcaagaagatctactgagatcgtgtgctccagggatc 501  
DB 301 atactgtgtgctgagatcactgtcaagaagatctactgagatcgtgtgctccagggatc 360  
QY 502 ggtaaacacacactggtggtttcaactagatcaatctgcaaatcccggtggaatggt 561  
DB 361 ggtaaacacacactggtggtttcaactagatcaatctgcaaatcccggtggaatggt 420  
QY 562 ggcctgtgtggaagacagttatataagatacagagggcagttcatgttgaacgtgtc 621  
DB 421 ggcctgtgtggaagacagttatataagatacagagggcagttcatgttgaacgtgtc 480  
QY 622 taccagatgtctgaaagggtgtatagggagacactgtgagacattccgacagccatgag 681  
DB 481 taccagatgtctgaaagggtgtatagggagacactgtgagacattccgacagccatgag 540  
QY 682 aagctctctgttccaaaacaaatcagcgtgtgagcgtttctgtggcgatatactatc 741  
DB 541 aagctctctgttccaaaacaaatcagcgtgtgagcgtttctgtggcgatatactatc 600  
QY 742 ttcggatatacgtgtacacacgaacaaattgtagtcataaactatgagagatcctc 801  
DB 601 ttcggatatacgtgtacacacgaacaaattgtagtcataaactatgagagatcctc 660  
QY 802 agagagacataaagatgtgctgataatgataatgataatgataatgataatgataatg 861  
DB 661 agagagacataaagatgtgctgataatgataatgataatgataatgataatgataatg 720  
QY 862 gattttgagatctgtgacatgtaggacagagatgagcgaatgagatgataatgagatg 921  
DB 721 gattttgagatctgtgacatgtaggacagagatgagcgaatgagatgataatgagatg 780  
QY 922 aagatgtcaaaacacataactgtgacgtgtgtgtgtgtgaaacagatcaactaaatt 981  
DB 781 aagatgtcaaaacacataactgtgacgtgtgtgtgtgtgaaacagatcaactaaatt 840  
QY 982 acagaaaggatcattcaatgactctgtctagtgtagcagcgtgtccactatgacg 1041  
DB 841 acagaaaggatcattcaatgactctgtctagtgtagcagcgtgtccactatgacg 900  
QY 1042 aaccggttatcttcacgtgacgtgagatggaacgaacagatgacacattgataagttct 1101  
DB 901 aaccggttatcttcacgtgacgtgagatggaacgaacagatgacacattgataagttct 960  
QY 1102 tcaattccagtagcctcagcacacgtatgagtgagacaggaagattagagatgtgt 1161  
DB 961 tcaattccagtagcctcagcacacgtatgagtgagacaggaagattagagatgtgt 1020  
QY 1162 agctcaaacacacagcagacccgagtagcgtatgcaattctgtgtcaagcattgtatgt 1221  
DB 1021 agctcaaacacacagcagacccgagtagcgtatgcaattctgtgtcaagcattgtatgt 1080  
QY 1222 ccactacgctctgacgcttcttcgcacatgagatcttttgacagtgaggtgagacgtg 1281  
DB 1081 ccactacgctctgacgcttcttcgcacatgagatcttttgacagtgaggtgagacgtg 1140  
QY 1282 agaatagtacacatttgtgtatctcagttgtctgtgtgcgtgtgtaacaaacactt 1341  
DB 1141 agaatagtacacatttgtgtatctcagttgtctgtgtgtgcgtgtgtaacaaacactt 1200  
QY 1342 aagagagaagtaataatacaacagacagcgtatataatggttttgaatcgaatctgac 1401  
DB 1201 aagagagaagtaataatacaacagacagcgtatataatggttttgaatcgaatctgac 1260  
QY 1402 ccattgtacattcagtaagcctataatagcgtgcaaaaaaataaaaaaataaaaaa 1461  
DB 1261 ccattgtacattcagtaagcctataatagcgtgcaaaaaaataaaaaaataaaaaa 1320  
QY 1462 aaaaaaaaaaaaaa 1474



Db 1074 ccactacgctccctgcagcgttctcttcgcacatgactcttttgacatagtagagtgagacg 1133

Qy 1282 agaatagtaeccatttgttgatcttcagtgttcgttgcgcgttgcgtacccaacacact 1341

Db 1134 agaatagtaeccat-----ttgatcttcagttgtcttgcgcgttggtctacaacacact 1189

Qy 1342 aagagagaataataacaagaacagcgctaatatagtttcttgcatactgaacatctg 1401

Db 1190 aagagagaagtaataacaagaacagcgctaatatagtttcttgcatactgaacatctg 1249

Qy 1402 ccactgtaacttcagttaagaagcccatatctgcgcgcaaaa 1441

Db 1250 ccactgtaacttcagttaagaagcccatatctgcgcgcatata 1289



Best Local Similarity 1.18; Pred. No. 1e-13;  
Matches 8; Conservative 439; Mismatches 302; Indels 0; Gaps 0;

QY	726	ggcgacatactctctccgataatgacgttacaccgaacaattgcagtcataataca	785
Db	790	GGCGATTTTCTCTCCGATAATGACGTTACACCGAACAATTGCAGTCATAATA	731
QY	786	catgagaagatctctcgaagagacataaagatgctgataatgatagtgatc	845
Db	730	WW	671
QY	846	ttccacctcgacaagatttgaagatctgcaactgagaccagagctaaagatt	905
Db	670	WW	611
QY	906	alcattgaatgaatgaagatgtcaagacataactgtgcaattgtcttgaaca	965
Db	610	WW	551
QY	966	atcactactaatttacagaaggctcatcttaattgactctgtctagtgacagctg	1025
Db	550	WW	491
QY	1026	gtccacctatgcacgaacggctgattctgcactggaatgggaacgaatacgcaca	1085
Db	490	WW	431
QY	1086	tcttgataagtcctctcactccagtgcctcgcacgcgtatgcagtgcacagcaag	1145
Db	430	WW	371
QY	1146	gattagagatgctgtgaagctcaaacacaaagcgagccgagtaacgtacgtctgtg	1205
Db	370	WW	311
QY	1206	tcaagcaactgtatgtcactaagctcctgcagcttcttcgcagatgacttggact	1265
Db	310	WW	251
QY	1266	agtgagtgagacgtgagaatgataccatttgttgcattcgcagttctgtgcgtg	1325
Db	250	WW	191
QY	1326	gctaccacaaccttaagagagaataataacaacagacagctaatatagtttg	1385
Db	190	WW	131
QY	1386	tatcgtaacatctggcccatcgtaactcagtaagcgtataatgacgggaacaaaaa	1445
Db	130	WW	71
QY	1446	aaaaaaaaaaaaaaaaaaaaaaaaa	1474
Db	70	WW	42
RESULT 7			
ID AAF58257/c			
AAF58257 standard; DNA; 936 BP.			
AAF58257;			
24-APR-2001 (first entry)			
Oligonucleotide D1954.			
Electron-transfer group; ETM; mismatch; genotyping;			
gene expression; ss.			
Synthetic.			
W0200107665-A2.			



[illegible]

```

QY 1086 tcttgataagtcctcctcacttcagtagcctcagcagcgtatgcagtgacagcaag 1145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 430 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 371
QY 1146 gattagagatgctgagagctaaaccacagcgagccgagtaacgtacgtcttggtg 1205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 370 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 311
QY 1206 tcaagcactgtatgtcactacgtcctcgcagcttccttcgcagatgacttttgact 1265
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 310 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 251
QY 1266 agtgagtgagagctggaagaatagtaaccttctgttgatcttcagtgctgttcgcttg 1325
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 250 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 191
QY 1326 gctaccacacacttaagagaagtaatacaacagacagctaatatagtgcttg 1385
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 190 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 131
QY 1386 tatctgaacatctgcccacgtacatcagtaagcctataatagcgcgcaaaaaaa 1445
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 130 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 71
QY 1446 aaaaaaaaaaaaaaaaaaaaaa 1474
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 70 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 42

```

## RESULT 11

AAF58252  
ID AAF58252 standard; DNA; 936 BP.

AC AAF58252;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1835.

XX Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

OS Synthetic.

XX WO200107665-A2.

PN 01-FEB-2001.

XX 26-JUL-2000; 2000WO-US20476.

PF 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX UmeK RM;

PI WPI; 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in

PT hybridization assays, e.g. for genotyping, allowing repeat analyses on

PS a single surface

XX Example 6; Page 127; 159pp; English.

CC The present invention relates to a composition comprising two nucleic

XX acids each containing an electron-transfer group (ETM) having

CC different redox potentials. The invention is used for electronic

XX detection of nucleic acids, especially of substitutions (mismatches)

CC and single-nucleotide polymorphisms, e.g. for genotyping,

XX monitoring gene expression.

SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 6.1%; Score 89.8; DB 22; Length 936;

Best Local Similarity 0.98; Pred. No. 1.3e-13;

Matches 7; Conservative 454; Mismatches 316; Indels 0; Gaps 0;

```

QY 698 aaaaacattacagcctgagcgttcttcgagatctcattacttcgagatcagct 757
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 9 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 68
QY 758 acacgacacaattgcagtcataactacatgagaagttcctcagagacataaagatg 817
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 69 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 128
QY 818 tgcgtatagttatattgtaagtgtaacttccacttcgacagaagtttgaaagatcgg 877
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 129 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 188
QY 878 cactgaagaccagagtgtaagtgatlatcattgaagtaatgaagattgcaagaacat 937
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 189 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 248
QY 938 ataactgagcagtgcttgcttgtaaccaagtcacttaattacagaagggtcattc 997
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 249 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 308
QY 998 aattgactctgctcctagtgagacagctggtccacatcagcagacggttattctgc 1057
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 309 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 368
QY 1058 actggaatgggaacgacgaacatcagcacatctgataagtcctcctcactcagtagcct 1117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 369 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 428
QY 1118 cagcacgctatgcagtgacgagcgaaggatagagatgctgtagctaaaccacagc 1177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 429 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 488
QY 1178 gagccgagtaacgtagactcttggtgtaagcaactgtatgtccactacgtctcgtca 1237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 489 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 548
QY 1238 gcttctcgccatgagatccttgagactgagagtgagagactggaagatgacatctt 1297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 549 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 608
QY 1298 gttgattcctcagtgcttggtggtggtggtggtggtggtggtggtggtggtggtggt 1357
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 609 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 668
QY 1358 caacagacagagctaatatagtggttgatctgacatctggccatcgtagctcagtg 1417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 669 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 728
QY 1418 aaagcctataatagcggtgcaaaaaaa 1474
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 729 wmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 785

```

## RESULT 12

AAF58254  
ID AAF58254 standard; DNA; 936 BP.

AC AAF58254;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1875.

XX Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

XX

```
OS Synthetic.
XX
XX MO200107665-A2.
XX
XX 01-FEB-2001.
XX
XX 26-JUL-2000; 2000MO-US20476.
XX
XX 26-JUL-1999; 99US-0145695.
XX
XX 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX umek RM;
XX
XX WPI; 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface.
XX
XX Example 6; Page 127; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX
XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;
XX
Query Match 6.1%; Score 89.8; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. No. 1.3e-13;
Matches 7; Conservative 454; Mismatches 316; Indels 0; Gaps 0;
XX
QY 698 aaaaacatacagcctgagcgttccctgagcgtatcttacttcgcgagatcgagt 757
DB 9 www. .... 68
QY 758 acacgacaacattgagcacaatacagagagagagagagagagagagagagatg 817
DB 69 www. .... 128
QY 818 tgcgtatagttattatgagtggttacttccacttcgcgacaagatttgaagatcgg 877
DB 129 www. .... 188
QY 878 cactgagagcagagtgctgaagtgaattgaattgaattgaattgaattgaattgaat 937
DB 189 www. .... 248
QY 938 ataacttgagcagtgcttggttgacacagtaacttaacttaacttaacttaacttaact 997
DB 249 www. .... 308
QY 998 aatgactctgtctagtgagagagtggtccactcagcagacgagtggtgattcgc 1057
DB 309 www. .... 368
QY 1058 actggaatggagacagacagacacacacacacacacacacacacacacacacacac 1117
DB 369 www. .... 428
QY 1118 cagcagcgtatgacgtgacgagcagaagagattgagatgctgagcctcaaaccaagc 1177
DB 429 www. .... 488
QY 1178 gaggccagtaagctagacattctgtgtcaagcattgatttcaactacgctcctgca 1237
DB 489 www. .... 548
```

```
QY 1238 gcttcttcgcacatgagcatttggactagtgagtgagagagagagagagagagag 1297
DB 549 www. .... 608
QY 1298 gtgattctcagtgcttggttgccgttgctaccacacacacacacacacacacacac 1357
DB 609 www. .... 668
QY 1358 caacagacagcagcacaatagtggttgatctgaacatctgcccacatcgatcagtt 1417
DB 669 www. .... 728
QY 1418 aaagcctataatagcgggcaaaaaaaacacacacacacacacacacacacacacac 1474
DB 729 www. .... 785
```

```
RESULT 13
AAF58257
ID AAF58257 standard; DNA; 936 BP.
XX
XX AAF58257;
XX
XX 24-APR-2001 (first entry)
XX
XX Oligonucleotide D1954.
XX
XX Electron-transfer group; ETM; mismatch; genotyping;
XX gene expression; ss.
XX
XX Synthetic.
XX
XX MO200107665-A2.
XX
XX 01-FEB-2001.
XX
XX 26-JUL-2000; 2000MO-US20476.
XX
XX 26-JUL-1999; 99US-0145695.
XX
XX 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX umek RM;
XX
XX WPI; 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface.
XX
XX Example 6; Page 127; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX
XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;
XX
Query Match 6.1%; Score 89.8; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. No. 1.3e-13;
Matches 7; Conservative 454; Mismatches 316; Indels 0; Gaps 0;
XX
QY 698 aaaaacatacagcctgagcgttccctgagcgtatcttacttcgcgagatcgagt 757
DB 9 www. .... 68
QY 758 acacgacaacattgagcacaatacagagagagagagagagagagagagagagatg 817
DB 69 www. .... 128
```

```

PI Unseq RM;
DR WP1: 2001-159728/16.
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface -
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 6.1%; Score 89.8; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. No. 1,3e-13;
Matches 7; Conservative 454; Mismatches 316; Indels 0; Gaps 0

QY 698 aaaacaattacagcgtgacgcttccctggcggatatactattacttcgagatgcagt 757
DB 9 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 68
QY 758 acacgcacaacatgcagtcataactaactatgtagagatccctcagagacataagatg 817
DB 69 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 128
QY 818 tgcgtatagttatattgtagttagtctaacttccacttcgacaagaattggaatctgg 877
DB 129 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 188
QY 878 cactgagagaccagagtgctaatgtagtatacttgaagttaatgaagatgcaagacat 937
DB 189 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 248
QY 938 ataactggcagttgctctgttgaaacaaagtcactaattacaagaaggatcttc 997
DB 249 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 308
QY 998 aattgactcttgcctcagtgacagctggtccactatgcagcaacggttgattctgc 1057
DB 309 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 368
QY 1058 actggaatggaagcaacagatcacacatctgataagtccttcacttcagtagcct 1117
DB 369 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 428
QY 1118 cagcacgcgtatgcagtgcagcagcgaagagattagagatgctgtgactcaaccacaag 1177
DB 429 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 488
QY 1178 gaggccgagtaacgtagcaactctgtgttcaagcactgttatgctcactacgtctgca 1237
DB 489 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 548
QY 1238 gcttctcttcgcacatgacatcttggactagtgaggtgagactgagaatgtacatctt 1297
DB 549 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 608
QY 1298 gttagttctcagttgcttggcogtltgctcaccaaccaaccttaagagagaatgaata 1357
DB 609 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 668
QY 1358 caacagacagcgttaataagttgttgcataacatctgcgcacatgacatcgaatcga 1417
DB 669 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 728
QY 1418 aaagcctataatagcggcaaaaaaanaaaaaaanaaaaaaanaaaaaaanaaaaaa 1474

```



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2001, 16:27:27 ; Search time 12.27 Seconds

(Without alignments)  
493.363 Million cell updates/sec

Title: US-09-537-654-2

Perfect score: 1530  
1 MODGSGNRNGPOKRYVSGAQ.....VTGKGRDAVSSNHRKARVT 294

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued\_Patents\_AA:\*  
2: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	323	21.1	350	4	US-09-157-603-1
2	323	21.1	350	4	US-09-587-436-1
3	313	20.5	350	4	US-09-157-603-3
4	313	20.5	350	4	US-09-587-436-3
5	281.5	18.4	339	3	US-08-758-280-2
6	281.5	18.4	339	3	US-08-664-614A-2
7	280.5	18.3	339	3	US-08-758-280-1
8	280.5	18.3	339	3	US-08-664-614A-1
9	126	8.2	781	2	US-09-114-637-2
10	126	8.2	781	2	US-09-114-637-2
11	107	7.0	451	3	US-08-444-189-8
12	107	7.0	451	3	US-08-444-189-8
13	107	7.0	451	3	US-08-444-189-8
14	101	6.6	273	1	US-08-215-928A-2
15	101	6.6	273	1	US-08-446-920-2
16	101	6.6	347	1	US-08-446-920-10
17	96	6.3	388	1	US-08-261-110A-4
18	96	6.3	438	1	US-08-261-110A-2
19	93.5	6.1	340	1	US-08-097-831-2
20	91.5	6.0	429	2	US-08-310-912A-109
21	91.5	6.0	429	2	PCT-US95-04589-109
22	91.5	6.0	1824	2	US-08-680-327-3
23	91.5	6.0	1824	2	US-09-228-246-2
24	87	5.7	353	1	US-08-097-831-6
25	86	5.6	432	2	US-08-644-034A-2
26	86	5.6	450	4	US-09-191-879-2
27	84.5	5.5	796	4	US-08-868-699A-2

28	84	5.5	2713	5	PCT-US96-01735-1	Sequence 1, Appl1
29	81.5	5.3	426	2	US-08-644-034A-3	Sequence 3, Appl1
30	78.5	5.1	186	1	US-08-077-256-3	Sequence 3, Appl1
31	78.5	5.1	186	1	US-08-259-672-3	Sequence 3, Appl1
32	78.5	5.1	186	1	US-08-459-351-3	Sequence 3, Appl1
33	78.5	5.1	186	1	US-08-460-533-3	Sequence 3, Appl1
34	78.5	5.1	186	3	PCT-US94-06654-3	Sequence 3, Appl1
35	78.5	5.1	189	5	US-09-209-666-11	Sequence 11, Appl1
36	78.5	5.1	189	3	US-09-356-952-1	Sequence 1, Appl1
37	78	5.1	459	2	US-08-472-172-2	Sequence 2, Appl1
38	76	5.0	1112	2	US-08-714-402-2	Sequence 2, Appl1
39	74.5	4.9	189	2	US-08-429-964-82	Sequence 82, Appl1
40	74	4.8	379	3	US-08-840-204-3	Sequence 3, Appl1
41	74	4.8	390	1	US-08-121-714-6	Sequence 6, Appl1
42	74	4.8	390	1	US-08-477-108A-6	Sequence 6, Appl1
43	74	4.8	390	2	US-08-477-112-6	Sequence 6, Appl1
44	74	4.8	390	5	PCT-US93-08322-6	Sequence 6, Appl1
45	74	4.8	399	1	US-08-565-386-16	Sequence 16, Appl1

#### ALIGNMENTS

```
RESULT 1
US-09-157-603-1
; Sequence 1, Application US/09157603
; Patent No. 6174694
; GENERAL INFORMATION:
; APPLICANT: Havre, Pamela A.
; APPLICANT: Rice, Michael C.
; APPLICANT: Holloman, William K.
; APPLICANT: Kmiec, Eric B.
; TITLE OF INVENTION: REC2 Kinase
; FILE REFERENCE: 7991-034-999
; CURRENT APPLICATION NUMBER: US/09/157,603
; NUMBER OF SEQ ID NOS: 8
; CURRENT FILING DATE: 1998-09-21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-157-603-1

Query Match          21.1%; Score 323; DB 4; Length 350;
Best Local Similarity 32.2%; Pred. No. 1.6e-29;
Matches 92; Conservative 44; Mismatches 100; Indels 50; Gaps 7;

QY 27 SPDLSSQKHITTTGSGDINDITLGGICKEVTEIGVGVGKTLGTLQIAINVOIPVECGGL 86
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 75 SADFSAPFLSTLISALDELHGVAGCSLTETITGPPGCKTKTFCIMMSTLATLPINMGL 134

QY 87 GGAAYITDEGSPFMRVYQIAEGCIRDTLEHPP--HSHEKSSSVOKOLOPERLP-ADI 142
   | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 135 EGAVVYIDTESAFSABRLVEIAE-----SRPRYFTEELTLTSSRVHLYREITCDE 187

QY 143 YYRRCISTYEQIAVITVMKFLREHKDVRIIVDSYTFHFRODFE-----DLALFTRVLS 197
   | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 188 VLORITSLEETI-----SKIKLVIDSVAIVKREFDQLOGNKKERKFLTA 236

QY 198 GSILKLMKIAKTYNLAVLLNVOYTKF-----TEGSPOTLTL 235
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 237 REASSLKLYALEERSIVILLNQTTLSGALASQADLVSPADLSLSEGTSSSCVIAL 296

QY 236 GDSWSHCNRLILHNGNERYAHL-DKSPSLPVASAPAVYTGKI 280
   | : | | | | | : | : | : | : | : | : | : | : | : | : | : |
DB 297 GNTWSHVNTRLIOTYLDSEKROIILAKSPILAFPTSFVYTIKEEGH 342

RESULT 2
US-09-587-436-1
; Sequence 1, Application US/09587436
```



Db 341 GL--VLOGHER 349

RESULT 5  
US-08-758-280-2

Sequence 2, Application US/08758280

Patent No. 6037125

GENERAL INFORMATION:

APPLICANT: Hasly, Paul

TITLE OF INVENTION: DISRUPTION OF THE MAMMALIAN Rad51

TITLE OF INVENTION: PROTEIN AND DISRUPTION OF PROTEINS THAT ASSOCIATE WITH

TITLE OF INVENTION: MAMMALIAN Rad51 FOR HINDERING CELL PROLIFERATION AND/OR

TITLE OF INVENTION: VIABILITY OF PROLIFERATING CELLS

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/758,280

FILING DATE: Concurrently herewith

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Halluin, Albert P.

REGISTRATION NUMBER: 25,227

REFERENCE/DOCKET NUMBER: 8535-013-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 854-3660

TELEFAX: (415) 854-3694

TELEX: 66141

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 339 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-758-280-2

Query Match 18.4%; Score 281.5; DB 3; Length 339;

Best Local Similarity 30.4%; Pred. No. 1.2e-24;

Matches 80; Conservative 46; Mismatches 96; Indels 41; Gaps 7;

35 ITTGGDNDLIDGGIHCKEVTEIGVGVGKTOGLQIALINVOIYVECGGAGKAVYID 94

102 ITTGGKELDKLGGIETGISTEMGEFRGTQICHTLAVTQQLPDRGGGKAMVYID 161

95 TEGSFWEVRYOIAE-----GCIKDILHFPHSHEKSSSVOKOLOPERFLADYIFRIS 149

162 TEGTFREPRERLLAVAEKGLSG--SDVLDNVAVARAFNTDQTL----- 204

150 YTEQIAVINMEKFLREHKDVRIIVDSVTFHPRDPE--DLALRTVLGSLKLMKI 206

205 -----YQASAMWESRYALLIVDSATALYRTDYSGRGELSRQMHARFLRL 255

207 AKTYNLAVLLNQVTTFTGSGFQTL-----ALGDSWSHSCNRLILH-WNGNERVAHL 260

256 ADEFGAVAVITNOVVAQ-VDGAAFAADPKPKIGNIIAHSTRLYLKRGGRGTRICKI 314

261 DKSPSLPVASAPYAVTGKIRDA 283

Db 315 YDSPCLPEAEAMFAINADGVDA 337

RESULT 6

US-08-964-614A-2

Sequence 2, Application US/08964614A

Patent No. 6057104

GENERAL INFORMATION:

APPLICANT: Hasly, Paul

TITLE OF INVENTION: DISRUPTION OF THE MAMMALIAN

TITLE OF INVENTION: Rad51 PROTEIN AND DISRUPTION OF PROTEINS THAT ASSOCIATE

TITLE OF INVENTION: WITH MAMMALIAN Rad51 FOR HINDERING CELL PROLIFERATION

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds, LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036-2811

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FASTSEQ for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/964,614A

FILING DATE: 05-NOV-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/758,280

FILING DATE: 05-NOV-1996

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 8535-0019-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-493-4935

TELEFAX: 650-493-5556

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 339 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 6057104e

US-08-964-614A-2

Query Match 18.4%; Score 281.5; DB 3; Length 339;

Best Local Similarity 30.4%; Pred. No. 1.2e-24;

Matches 80; Conservative 46; Mismatches 96; Indels 41; Gaps 7;

35 ITTGGDNDLIDGGIHCKEVTEIGVGVGKTOGLQIALINVOIYVECGGAGKAVYID 94

102 ITTGGKELDKLGGIETGISTEMGEFRGTQICHTLAVTQQLPDRGGGKAMVYID 161

95 TEGSFWEVRYOIAE-----GCIKDILHFPHSHEKSSSVOKOLOPERFLADYIFRIS 149

162 TEGTFREPRERLLAVAEKGLSG--SDVLDNVAVARAFNTDQTL----- 204

150 YTEQIAVINMEKFLREHKDVRIIVDSVTFHPRDPE--DLALRTVLGSLKLMKI 206

205 -----YQASAMWESRYALLIVDSATALYRTDYSGRGELSRQMHARFLRL 255

207 AKTYNLAVLLNQVTTFTGSGFQTL-----ALGDSWSHSCNRLILH-WNGNERVAHL 260

256 ADEFGAVAVITNOVVAQ-VDGAAFAADPKPKIGNIIAHSTRLYLKRGGRGTRICKI 314

261 DKSPSLPVASAPYAVTGKIRDA 283

315 YDSPCLPEAEAMFAINADGVDA 337



Tue Oct 2 11:16:27 2001

us-09-537-654-2.ra1

Page 5

RESULT 9

US-08-373-134D-2

Sequence 2, Application US/08373134D

Patent No. 5780296

GENERAL INFORMATION:

APPLICANT: Kmiec, Eric

APPLICANT: Holloman, William

TITLE OF INVENTION: COMPOSITIONS AND METHODS TO PROMOTE

TITLE OF INVENTION: HOMOLOGOUS RECOMBINATION IN EUKARYOTIC CELLS AND ORGANISMS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/373,134D

FILING DATE: January 17, 1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Fiebel, Thomas E.

REGISTRATION NUMBER: 29,258

REFERENCE/DOCKET NUMBER: 7991-007

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 781 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-373-134D-2

Query Match 8.2%; Score 126; DB 1; Length 781;

Best Local Similarity 20.5%; Pred. No. 9.4e-06; Mismatches 74; Indels 114; Gaps 13;

Matches 62; Conservative 52; Mismatches 74; Indels 114; Gaps 13;

Query 28 DELS---OKHT-TGSGDLNDILGGGICKEVTEIGVPGVCKTGLGLAIN----- 76

DB 215 DVLISGRORHVFSSGSELDLGGVRSVAVLTVGVSGSGSKTQMAIOVCTYAAALGLVP 274

QY 77 -----VOIPEVCG-----GIGGKRAVYID 94

DB 275 LSQADDHDKGNNTFQSRFFVADPIHASTKDDTLSDILQSYGMEPSIGSHRGK--ACYIT 332

QY 95 TEGS---FMERYVOIAEGCIRDLIEH-FPHSHEKSSVQ-----KOLQPER 137

DB 333 SGGERAAHSIVRALELASFALINERFDRIYVPCDPTQSSQDADGRDALLAKAQLGRG 392

QY 138 FLADIIYFRICSYTEQIAVINMEKFLR-----EKKYVRVIITDSVT 179

DB 393 ALANLHI--AC-----VADVEALEHALKYSPLGLIRRLMSKROSGVSRIGVAVVNDLP 445

QY 180 FHFRRQ-----DFEDLARTVLSGLSLKMKIA-----KTYNLAVALVLLNOVTT 222

DB 446 ALFQDDQAAASDIDSLFORSKMLVEITADALKRISAVQWRGASDGGSSAGRAVLYLVNHVSD 505

QY 223 KF 224

DB 506 AF 507

RESULT 10

US-09-114-637-2

Sequence 2, Application US/09114637

Patent No. 5945339

GENERAL INFORMATION:

APPLICANT: Kmiec, Eric

APPLICANT: Holloman, William

TITLE OF INVENTION: COMPOSITIONS AND METHODS TO PROMOTE

TITLE OF INVENTION: HOMOLOGOUS RECOMBINATION IN EUKARYOTIC CELLS AND ORGANISMS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/114,637

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/373,134

FILING DATE: January 17, 1995

ATTORNEY/AGENT INFORMATION:

NAME: Fiebel, Thomas E.

REGISTRATION NUMBER: 29,258

REFERENCE/DOCKET NUMBER: 7991-007

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 781 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-114-637-2

Query Match 8.2%; Score 126; DB 2; Length 781;

Best Local Similarity 20.5%; Pred. No. 9.4e-06; Mismatches 74; Indels 114; Gaps 13;

Matches 62; Conservative 52; Mismatches 74; Indels 114; Gaps 13;

Query 28 DELS---OKHT-TGSGDLNDILGGGICKEVTEIGVPGVCKTGLGLAIN----- 76

DB 215 DVLISGRORHVFSSGSELDLGGVRSVAVLTVGVSGSGSKTQMAIOVCTYAAALGLVP 274

QY 77 -----VOIPEVCG-----GIGGKRAVYID 94

DB 275 LSQADDHDKGNNTFQSRFFVADPIHASTKDDTLSDILQSYGMEPSIGSHRGK--ACYIT 332

QY 95 TEGS---FMERYVOIAEGCIRDLIEH-FPHSHEKSSVQ-----KOLQPER 137

DB 333 SGGERAAHSIVRALELASFALINERFDRIYVPCDPTQSSQDADGRDALLAKAQLGRG 392

QY 138 FLADIIYFRICSYTEQIAVINMEKFLR-----EKKYVRVIITDSVT 179

DB 393 ALANLHI--AC-----VADVEALEHALKYSPLGLIRRLMSKROSGVSRIGVAVVNDLP 445

QY 180 FHFRRQ-----DFEDLARTVLSGLSLKMKIA-----KTYNLAVALVLLNOVTT 222

DB 446 ALFQDDQAAASDIDSLFORSKMLVEITADALKRISAVQWRGASDGGSSAGRAVLYLVNHVSD 505

QY 223 KF 224

DB 506 AF 507

RESULT 11

US-08-969-644-8

; Sequence 8, Application US/08969644

; Patent No. 6096519

; GENERAL INFORMATION:

; APPLICANT: Rattl, Giulio

; APPLICANT: Comanducci, Maurizio

; APPLICANT: Tecce, Mario F.

; APPLICANT: Giuliani, Marzia M.

; TITLE OF INVENTION: PCTD PLASMID ISOLATED FROM CHLAMYDIA

; TITLE OF INVENTION: TRACHOMATIS SEROTYPE D, ITS GENES AND PROTEINS ENCODED BY

; TITLE OF INVENTION: THEM: RECOMBINANT PLASMIDS FOR THE EXPRESSION OF SAID

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BIRCH, STEWART, KOLASCH &amp; BIRCH

; STREET: 301 N. Washington Street

; CITY: Falls Church

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22046-0747

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; FILING DATE: 13-NOV-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/467,152

; FILING DATE: US/07/661,820

; APPLICATION NUMBER: IT MI 91A000314

; FILING DATE: 07-FEB-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Svensson, Leonard R.

; REGISTRATION NUMBER: 30,330

; REFERENCE/DOCKET NUMBER: 1267-202P

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-241-1300

; TELEFAX: 703-241-2848

; TELEEX: 248345

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 451 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-969-644-8

; US-08-969-644-8

; US-08-969-644-8

; US-08-969-644-8

; US-08-969-644-8

; US-08-969-644-8

; US-08-969-644-8

; US-08-969-644-8

; US-08-969-644-8

; US-08-969-644-8

; US-08-969-644-8

; US-08-969-644-8

; US-08-969-644-8

; US-08-969-644-8

; US-08-969-644-8

; US-08-969-644-8

Db 367 -----NIPVLSLSRK 379

RESULT 12

US-08-444-189-8

; Sequence 8, Application US/08444189

; Patent No. 6110705

; GENERAL INFORMATION:

; APPLICANT: Rattl, Giulio

; APPLICANT: Comanducci, Maurizio

; APPLICANT: Tecce, Mario F.

; APPLICANT: Giuliani, Marzia M.

; TITLE OF INVENTION: PCTD PLASMID ISOLATED FROM CHLAMYDIA

; TITLE OF INVENTION: TRACHOMATIS SEROTYPE D, ITS GENES AND PROTEINS ENCODED BY

; TITLE OF INVENTION: THEM: RECOMBINANT PLASMIDS FOR THE EXPRESSION OF SAID

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BIRCH, STEWART, KOLASCH &amp; BIRCH

; STREET: 301 N. Washington Street

; CITY: Falls Church

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22046-0747

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; FILING DATE: US/08/444,189

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/180,528

; FILING DATE: US/07/991,512

; APPLICATION NUMBER: IT MI 91A000314

; FILING DATE: 07-FEB-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Svensson, Leonard R.

; REGISTRATION NUMBER: 30,330

; REFERENCE/DOCKET NUMBER: 1267-202P

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-241-1300

; TELEFAX: 703-241-2848

; TELEEX: 248345

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 451 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-444-189-8

; US-08-444-189-8

; US-08-444-189-8

; US-08-444-189-8

; US-08-444-189-8

; US-08-444-189-8

; US-08-444-189-8

; US-08-444-189-8

; US-08-444-189-8

; US-08-444-189-8

; US-08-444-189-8

; US-08-444-189-8

; US-08-444-189-8

; US-08-444-189-8





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 04:00:14 ; Search time 1172.69 Seconds  
(without alignments)  
11881.662 Million cell updates/sec

Title: US-09-537-654-1  
Perfect score: 1474  
Sequence: 1 tcgaccacagctgcgcact.....aaaaaaaaaaaaaaaaaaaaa 1474

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_esthum29:\*  
64: em\_esthum30:\*  
65: em\_esthum31:\*  
66: em\_esthum32:\*  
67: em\_esthum33:\*  
68: em\_esthum34:\*  
69: em\_esthum35:\*  
70: em\_esthum36:\*  
71: em\_esthum37:\*  
72: em\_esthum38:\*  
73: em\_esthum39:\*  
74: em\_esthum40:\*  
75: em\_esthum41:\*  
76: em\_esthum42:\*  
77: em\_esthum43:\*  
78: em\_esthum44:\*  
79: em\_esthum45:\*  
80: em\_esthum46:\*  
81: em\_esthum47:\*  
82: em\_esthum48:\*  
83: em\_esthum49:\*  
84: em\_esthum50:\*  
85: em\_esthum51:\*  
86: em\_esthum52:\*  
87: em\_esthum53:\*  
88: em\_esthum54:\*  
89: em\_esthum55:\*  
90: em\_esthum56:\*  
91: em\_esthum57:\*  
92: em\_esthum58:\*  
93: em\_esthum59:\*  
94: em\_esthum60:\*  
95: em\_esthum61:\*  
96: em\_esthum62:\*  
97: em\_esthum63:\*  
98: em\_esthum64:\*  
99: em\_esthum65:\*  
100: em\_esthum66:\*  
101: em\_esthum67:\*  
102: em\_esthum68:\*  
103: em\_esthum69:\*  
104: em\_esthum70:\*  
105: em\_esthum71:\*  
106: em\_esthum72:\*  
107: em\_esthum73:\*  
108: em\_esthum74:\*  
109: em\_esthum75:\*  
110: em\_esthum76:\*  
111: em\_esthum77:\*  
112: em\_esthum78:\*  
113: em\_esthum79:\*  
114: em\_esthum80:\*  
115: em\_esthum81:\*  
116: em\_esthum82:\*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: em_gss_fun:*
182: em_gss_hum1:*
183: em_gss_hum2:*
184: em_gss_hum3:*
185: em_gss_hum4:*
186: em_gss_hum5:*
187: em_gss_hum6:*
188: em_gss_hum7:*
189: em_gss_hum8:*
190: em_gss_hum9:*
191: em_gss_inv1:*
192: em_gss_inv2:*
193: em_gss_inv3:*
194: em_gss_inv4:*
195: em_gss_inv5:*
196: em_gss_inv6:*
197: em_gss_inv7:*
198: em_gss_inv8:*
199: em_gss_inv9:*
200: em_gss_inv10:*
201: em_gss_inv11:*
202: em_gss_inv12:*
203: em_gss_inv13:*
204: em_gss_inv14:*
205: em_gss_inv15:*
206: em_gss_inv16:*
207: em_gss_inv17:*
208: em_gss_inv18:*
209: em_gss_inv19:*
210: em_gss_inv20:*
211: em_gss_inv21:*
212: em_gss_inv22:*
213: em_gss_inv23:*
214: em_gss_inv24:*
215: em_gss_inv25:*
216: em_gss_inv26:*
217: em_gss_inv27:*
218: em_gss_inv28:*
219: em_gss_inv29:*
220: em_gss_inv30:*
221: em_gss_inv31:*
222: em_gss_inv32:*
223: em_gss_inv33:*
224: em_gss_inv34:*
225: em_gss_inv35:*
226: em_gss_inv36:*
227: em_gss_inv37:*
228: em_gss_inv38:*
229: em_gss_inv39:*
230: em_gss_inv40:*
231: em_gss_inv41:*
232: em_gss_inv42:*
233: em_gss_inv43:*
234: em_gss_inv44:*
235: em_gss_inv45:*
236: em_gss_inv46:*
237: em_gss_inv47:*
238: em_gss_inv48:*
239: em_gss_inv49:*
240: em_gss_inv50:*
241: em_gss_inv51:*
242: em_gss_inv52:*
243: em_gss_inv53:*
244: em_gss_inv54:*
245: em_gss_inv55:*
246: em_gss_inv56:*
247: em_gss_inv57:*
248: em_gss_inv58:*
249: em_gss_inv59:*
250: em_gss_inv60:*
251: em_gss_inv61:*
252: em_gss_inv62:*
253: em_gss_inv63:*
254: em_gss_inv64:*
255: em_gss_inv65:*
256: em_gss_inv66:*
257: em_gss_inv67:*
258: em_gss_inv68:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length	ID
1	438	29.7	596	173 BG051275
2	382	25.9	543	155 BG605725
3	372.2	25.3	614	155 BG605648
4	367.2	24.9	603	136 BG517872
5	311	21.1	523	251 AZ218279
6	300	20.4	455	251 AZ2918278
7	252.8	17.2	481	173 BG505095
8	244.6	16.6	670	114 AM310460
9	177.6	12.0	606	113 AM203650
10	102.4	6.9	506	120 AW784884
11	100.4	6.8	550	154 BG471883
12	99.6	6.8	692	17 A1184177
13	98	6.6	658	167 BE390998
14	98	6.6	683	167 BE3907492
15	98	6.6	778	174 BG477028
16	95.6	6.5	676	154 BG477452
17	95	6.4	669	154 BG506121
18	92.4	6.3	899	168 BF672049
19	91.8	6.2	572	149 BF516288
20	88.8	6.0	634	234 AO848265
21	88.2	6.0	504	114 AM327505
22	85.2	5.8	460	148 BF448483
23	82.6	5.6	527	12 AA861390
24	80.2	5.4	468	31 AV629583
25	78.8	5.3	371	10 AA641847
26	77.4	5.3	545	174 BG431647
27	77.2	5.2	635	153 BG403188
28	77.2	5.2	870	146 BF240492
29	73.4	5.0	787	106 AL353302
30	72.4	4.9	535	154 BG521224
31	70	4.7	606	148 AM604548
32	66.8	4.5	744	165 BE281543
33	66.6	4.5	899	245 AZ544861
34	66.6	4.5	938	248 AZ667391
35	66	4.5	975	106 AL353348
36	65	4.4	900	247 AZ677092
37	65	4.4	925	245 AZ549236
38	64.4	4.4	647	136 BE492323
39	63.4	4.3	649	231 AO641696
40	61.6	4.2	481	10 AA707444
41	60.8	4.1	892	146 BF216102
42	60	4.1	984	219 CNS0055A
43	59.8	4.1	480	11 AA779979
44	59	4.0	489	226 AO254558
45	59	4.0	819	106 AL530472
ALIGNMENTS				
RESULT 1				
LOCUS BG051275	596 bp	RNA	EST	25-JAN-2001
DEFINITION FM1_54.D07.g1.A003	Floral-induced Meristem 1 (FM1)	Sorghum		
ACCESSION BG051275	propinquum cDNA, mRNA sequence.			
VERSION BG051275.1	GI:12504866			
KEYWORDS	EST.			
SOURCE	Sorghum propinquum.			
ORGANISM	Sorghum propinquum			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.			
AUTHORS	1 (bases 1 to 596) Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt			
TITLE	Floral-induced meristems			

RESULT 2  
 LOCUS BG605725 543 bp mRNA EST 17-APR-2001  
 DEFINITION RHIZ2\_80.G06.g1.A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA sequence.  
 ACCESSION BG605725 GI:13655696  
 VERSION BG605725.1  
 KEYWORDS EST.  
 SOURCE Sorghum propinquum.  
 ORGANISM Sorghum propinquum.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.  
 REFERENCE 1 (bases 1 to 543)  
 Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt, L. H.  
 AUTHORS An EST database from Sorghum: Sorghum propinquum rhizomes  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmp@pratt.uga.edu  
 Seg primer: PolyTMix  
 High quality sequence start: 39  
 High quality sequence stop: 542  
 POLYA-No.  
 FEATURES  
 source Location/Qualifiers  
 1..543  
 /organism="Sorghum propinquum"  
 /db\_xref="taxon:132711"  
 /clone\_lib="Rhizome2 (RHIZ2)"  
 /note="Organ: Rhizomes; Vector: pBluescript II from Lambda Zap II; Site\_1: XhoI; Site\_2: EcoRI; The library was made from poly-A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision."  
 BASE COUNT 152 a 122 c 119 g 150 t  
 ORIGIN  
 Query Match 25.9%; Score 382; DB 155; Length 543;  
 Best Local Similarity 91.6%; Pred. No. 1.9e-90;  
 Matches 416; Conservative 0; Mismatches 35; Indels 3; Gaps 1;

DB 361 GCTACAAACGACGACC---ATAGAGAAGTAGATGCAACGACGACTACTAGTGTG 417  
 QY 1386 tatctgaacatctgcccacatgacatcagtaa 1419  
 DB 418 TATCTGAACATGTGCGACATCTTGCATTCAGTAA 451  
 RESULT 3  
 LOCUS BG606048 614 bp mRNA EST 17-APR-2001  
 DEFINITION RHIZ2\_80.G06.b1.A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA sequence.  
 ACCESSION BG606048 GI:13656019  
 VERSION BG606048.1  
 KEYWORDS EST.  
 SOURCE Sorghum propinquum.  
 ORGANISM Sorghum propinquum.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.  
 REFERENCE 1 (bases 1 to 614)  
 Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt, L. H.  
 AUTHORS An EST database from Sorghum: Sorghum propinquum rhizomes  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmp@pratt.uga.edu  
 Seg primer: JEN REV  
 High quality sequence stop: 605  
 POLYA-No.  
 FEATURES  
 source Location/Qualifiers  
 1..614  
 /organism="Sorghum propinquum"  
 /db\_xref="taxon:132711"  
 /clone\_lib="Rhizome2 (RHIZ2)"  
 /note="Organ: Rhizomes; Vector: pBluescript II from Lambda Zap II; Site\_1: XhoI; Site\_2: EcoRI; The library was made from poly-A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision."  
 BASE COUNT 160 a 146 c 160 g 148 t  
 ORIGIN  
 Query Match 25.3%; Score 372.2; DB 155; Length 614;  
 Best Local Similarity 95.5%; Pred. No. 7.7e-88;  
 Matches 383; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
 QY 358 tcagagccaggaatgctggtgatgtctctgatagaactgtcaagaagaacatcaact 417  
 DB 214 TCTGGGCGCCAGATGCTGGGATGCTATCTGATGACGACGACACGAAACATCACT 273  
 QY 418 actggtctgtgtaacctcaatgacatactgtgtggtggaattcaacgcaagaattact 477  
 DB 274 ACTGTGCTGTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 333  
 QY 478 gagatcgtgtggtcccaagggttggtaaaactcaactggtggaattcaactgaatcaat 537  
 DB 334 GAGATGCTGTGCTGCCAGGCGTGGTAAACACTGCGGATTCACATCACTCAAT 393  
 QY 538 gtacaatcccaatgaaagt 597  
 DB 394 GTACAAATCCCACTGGAATATGATGATGATGATGATGATGATGATGATGATGATG 453  
 QY 598 ggcattcaatgtgaactgtcaccagatgtcgaagggtgtattaggaactactg 657  
 DB 454 GGAAGTTTATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 513

QY	568	gagacatttcgcgcacgcacatgaaagctccctctctgtcccaaacattacagccctgag	717
Db	514	GAGACATTTCGCACGCACACCATGAAAGTCTCCTCTCTGTGGCAAAAAACATTAACACCTTGAG	573
QY	718	cgtttcctgagcgagatcattcattactcccgatagcagttta	758
Db	574	CATTTCCTGGCGGACATCTATTACTTCGCAATATGACAGTTA	614
RESULT	4		
LOCUS	BS517872		
DEFINITION	BS517872	603 bp mRNA	EST
ACCESSION	WHE0804.H03_P06S	Wheat vernalized crown cDNA library	Trilicium
VERSION	aestium cDNA clone WHE0804_H03_P06,	mRNA sequence.	
KEYWORDS	BS517872		
SOURCE	BS517872.1	GI:9741902	
ORGANISM	EST.		
REFERENCE	bread wheat.		
AUTHORS	Trilicium aestivum		
TITLE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae		
JOURNAL	Trilicium		
COMMENT	1 (bases 1 to 603)		
	Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han		
	P.S., Hala,C.C., Kang,Y., Iazo,G.R., Miller,R., Rausch,C.J.,		
	Seaton,C.L. and Tong,J.C.		
	The structure and function of the expressed portion of the wheat		
	genomes - Vernalized crown cDNA library		
	Unpublished (2000)		
	Contact: Olin Anderson		
	US Department of Agriculture, Agriculture Research Service, Pacific		
	West Area, Western Regional Research Center		
	800 Buchanan Street, Albany, CA 94710, USA		
	Tel: 5105595773		
	Fax: 5105595818		
	Email: oanderson@wpr.usda.gov		
	Sequence have been trimmed to remove vector sequence and low		
	quality sequence with phred score less than 20		
	Seq primer: StrataGene SK primer.		
FEATURES	Location/Qualifiers		
source	1..603		
	/organism="Trilicium aestivum"		
	/cultivar="Chinese Spring"		
	/db_xref="taxon:4565"		
	/clone="WHE0804.H03_P06"		
	/clone_lib="Wheat vernalized crown cDNA library"		
	/tissue_type="Crown tissue of seedling"		
	/dev_stage="Five-week old seedling"		
	/lab_host="E. coli SOLR"		
	/note="Vector: lambda Uni-ZAP XR, excised phagemid;		
	Site1: EcoRI, Site2: XhoI; Seeds were germinated and		
	grown at 4 C for 5 weeks. The tissue, total RNA, and		
	poly(A) RNA were prepared, a cDNA library was made, and		
	the cDNA clones were in vivo excised to give pluescript		
	phagemids in the T7 Close lab (Choi, Close, Fenton) at		
	the University of California, Riverside. Plasmid DNA		
	preparations and DNA sequencing were performed in the OD		
	Anderson lab (all other authors)."		
BASE COUNT	175 a 136 c 158 g 134 t		
ORIGIN			
Query Match	24.9%; Score 367.2; DB 136; Length 603;		
Best Local Similarity	83.4%; Pred. No. 1,6e-86;		
Matches	417; Conservative 0; Mismatches 83; Indels 0; Gaps 0;		
QY	341	cacacagaagaagtaagcttcaagagcccaagaatgacctgagatattgtctctgtagcgtgt	400
Db	104	CCGACAGAGCTGCTCTCCACAGAGGCCCAAAAGCATGAGATGCTGCTGAGAGCAAT	163
QY	401	cacagaacacactactctgttctgtgtagacctaatbacaatttggtgcggagattc	460
Db	164	CACAGAAACACATCACTACTGCTGTGGGACCTGAACGCACTACTGCTGGAGGATTC	223

Y	461	acgcaagaagaactttagatcggctggccgtcccaagggttggfcaaaactcaactgggga	520
Y	461	acgcaagaagaactttagatcggctggccgtcccaagggttggfcaaaactcaactgggga	520
Db	224	ATTGTAAGAAGACTGACTGAGATAGGTGGGCTCCACGGAATTGGTAAACTCACTGGGGA	283
Y	521	ttcaactagacaatcatgtcacaaatccagtlggaatgtgtgtgctctgtgtggaaagcag	560
Db	284	TTTCAGCTAGCAATCAACGTCCTCAATTCGCCGTGACTATGTGTGGCTTGGCGGCAAGCA	343
Y	581	ttctatagatcacagagggcagtttccatggtgtgaacgtgtctaccagatgctggaaggt	640
Y	641	gtttttggagcagctacacggagcagcttccgcagacagcagagaagtcctctcttccaaa	700
Db	404	GCATCATGTGACATTAATGGAGTATTTTTCCTCATGCCACCATGACAAAGTCGATCTGGCCAG	463
Y	701	aacaattacagccttgagcgcttctcttggcggatcatctactctccgatalgtagtaca	760
Db	464	AAATCTGCAGCCCGAGAGTTTCCTGCGAGAACTACTACTTCGGAATATGACAGCTACA	523
Y	761	ccgacaacaattgcagtcataaactaaatgagaagtctctcagagagataaagtgc	820
Db	524	CTGAAACGATTTGGGGGATTAACATATCTGGAGAAATCTCTCGAGACATTAAGATGGA	583
Y	821	gtatagttatctatgtaagt 840	
Db	584	GAATAGTTATATTGATAGT 603	
RESULT	5		
LOCUS	AZ918279	523 bp	DNA
DEFINITION	1006003H03.x2	1006 - RescuedMu Grid G	zea mays genomic, DNA
ACCESSION	AZ918279		
VERSION	AZ918279.1	GI:13387563	
KEYWORDS	GSS.		
SOURCE	zea mays.		
ORGANISM	zea mays.		
REFERENCE			
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
TITLE	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC		
JOURNAL	clade: Panicoideae; Andropogoneae; Zea.		
COMMENT	1 (bases 1 to 523)		
	Maize genomic sequences found using engineered RescuedMu transposon		
	Unpublished (2001)		
	Contact: Walbot V		
	Department of Biological Sciences		
	Stanford University		
	855 California Ave, Palo Alto, CA 94304, USA		
	Tel: 650 723 2227		
	Fax: 650 725 8221		
	Email: walbot@stanford.edu		
	Possible ligation site or ends cut by 2 different endonucleases.		
	Reverse complemented post-ligation sequence from source sequence.		
	Plate: 1006003 row: H column: 03		
	Class: transposon-tagged.		
FEATURES			
SOURCE			
	1..523		
	/organism="zea mays"		
	/cultivar="mixed background W23/Al188/B73"		
	/db_xref="taxon:4577"		
	/clone_lib="1006 - RescuedMu Grid G"		
	/tissue_type="leaf"		
	/dev_stage="adult"		
	/lab_host="DH10B"		
	/note="Organ: leaf; Vector: RescuedMu (engineered from		
	phiRescript backbone); Site_1: BamHI; Site_2: BglII;		
	RescuedMu is a 4.9 kb, modified maize Mu transposon		
	designed to allow plasmid rescue from total genomic DNA.		
	Mu elements insert preferentially into transcription		
	units. For more information on RescuedMu, go to the web		

site 'www.zmdb.iastate.edu' and follow the links for 'RescueMu.' Grid G was grown at Stanford in 2000. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin.

BASE COUNT 109 a 140 c 139 g 135 t

ORIGIN

Query Match 21.1%; Score 311; DB 251; Length 523;  
Best Local Similarity 100.0%; Pred. No. 1.2e-71;  
Matches 311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 cacttgactccagctccactgtgcagctgtgtgtcccgagcccaaggcg 76  
DB 34 CACTTGACTCCAGCTCCACTGTGCGGAGTTCCTTGTCGCCGAGCCCAAGGCG 93  
QY 77 ggcgtgagccgagcccgagagacagcgcgagcagcagcagcagcagc 136  
DB 94 GCGGTGAGCCGAGAGCCGAGAGAGAGCGCGCGACTCCCTTAAGCGACGCGC 153  
QY 137 ggcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 196  
DB 154 GCGGTGAGCGAGAGCGGCTGCGGCGCGAGAGGCTGAGGCTGAGGCG 213  
QY 197 agatagagagtgagtgagtgagtgagtgagtgagtgagtgagtgag 256  
DB 214 AGATATGAGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGG 273  
QY 257 ctgcagctctctctctcccaatcagtaacacctctccaagtggcaatcgag 316  
DB 274 CTGCAGTCT 333  
QY 317 atcaatctggc 327  
DB 334 ATCATCTGCGC 344

RESULT 6  
A2918278 455 bp DNA GSS 20-MAR-2001  
LOCUS 1006003H03.x1 1006 - RescueMu Grid G Zea mays genomic, DNA  
DEFINITION  
Sequence.  
ACCESSION A2918278  
VERSION A2918278.1 GI:13387562  
KEYWORDS GSS.  
SOURCE Zea mays.  
ORGANISM Zea mays.

REFERENCE 1 (bases 1 to 455)  
AUTHORS Walbot,V.  
TITLE Maize genomic sequences found using engineered RescueMu transposon  
JOURNAL Unpublished (2001)  
COMMENT Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Possible ligation site of ends cut by 2 different endonucleases.  
Reverse complemented post-ligation sequence from source sequence.  
Plate: 1006003 row: H column: 03  
Class: transposon-tagged.

FEATURES  
source  
1..455  
Location/Qualifiers

Location/Qualifiers  
1..455  
/organism="Zea mays"  
/culturvar="mixed background W23/A188/B73"  
/db\_xref="taxon:4577"  
/clone\_id="1006 - RescueMu Grid G"

/tissue\_type="leaf"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: leaf. Vector: RescueMu (engineered from Bluescript backbone). Site: BamHI; Site\_2: BglII; RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'RescueMu.' Grid G was grown at Stanford in 2000. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

BASE COUNT 91 a 129 c 123 g 111 t 1 others

ORIGIN

Query Match 20.4%; Score 300; DB 251; Length 455;  
Best Local Similarity 99.7%; Pred. No. 9.8e-69;  
Matches 311; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 17 cacttgactccagctccactgtgcagctgtgtgtcccgagcccaaggcg 75  
DB 34 CACTTGACTCCAGCTCCACTGTGCGGAGTTCCTTGTCGCCGAGCCCAAGGCG 93  
QY 76 ggcgtgagccgagcccgagagacagcgcgagcagcagcagcagcagc 135  
DB 94 GCGGTGAGCCGAGAGCCGAGAGAGAGCGCGCGACTCCCTTAAGCGACGCGG 153  
QY 136 cgcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 195  
DB 154 CCGGTGAGCGAGAGCGGCTGCGGCGCGAGAGGCTGAGGCTGAGGCG 213  
QY 196 cagatagagagtgagtgagtgagtgagtgagtgagtgagtgagtgag 255  
DB 214 CAGATATGAGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGG 273  
QY 256 accgagctctctctcccaatcagtaacacctctccaagtggcaatcgag 315  
DB 274 ACTGCACTTCT 333  
QY 316 gatcaatctggc 327  
DB 334 GATCAATCTGCGC 345

RESULT 7  
BG050955 481 bp mRNA EST 25-JAN-2001  
LOCUS FMI\_54.D07.b1.A003 Floral-induced Meristem 1 (FMI) Sorghum  
DEFINITION  
Sorghum propinquum  
ACCESSION BG050955  
VERSION BG050955.1 GI:12504184  
KEYWORDS EST.  
SOURCE Sorghum propinquum.  
ORGANISM Sorghum propinquum.

REFERENCE 1 (bases 1 to 481)  
AUTHORS Cordomier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt L.H.  
TITLE An EST database from Sorghum: floral-induced meristems  
JOURNAL Unpublished (2000)  
COMMENT Contact: Cordomier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmpratt@uga.edu







RESULT 12  
 A1184177/c 692 bp mRNA EST 10-NOV-1998  
 LOCUS q14608.x1 Soares\_testis.NHT Homo sapiens cDNA clone IMAGE:1733094  
 DEFINITION 3' similar to TR:02144 022144 POTATIVE DNA REPAIR PROTEIN  
 ; contains Alu repetitive element; contains element MER22 repetitive  
 element ; mRNA sequence.

ACCESSION A1184177  
 VERSION A1184177.1 GI:3734815  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 692)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 Tumor Gene Index  
 Unpublished (1997)  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
 , Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.blo.illn.gov/bhrp/image/image.html](http://www.blo.illn.gov/bhrp/image/image.html)  
 Insert Length: 892 Std Error: 0.00  
 Seq primer: -400P from Gldco  
 High quality sequence stop: 444.  
 Location/Qualifiers  
 1. 692  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="1733094"  
 /clone\_lib="Soares\_testis.NHT"  
 /sex="male"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from Clontech Laboratories  
 , Inc., and primed with a Not I - oligo(dT) primer [5'  
 TGTTCACATCTGAAGTGGAGCGCGCCCAATTTTCTTTT 3']  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73D vector. Library  
 went through one round of normalization to Cots, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 195 a 163 c 153 g 181 t  
 ORIGIN

Query Match 6.8%; Score 99.6; DB 17; Length 692;  
 Best Local Similarity 60.9%; Pred. No. 1.4e-15;  
 Matches 162; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 392 atgagctgcacagaacacacactcactcgtctcgtcgaactcaatgacatactgtg 451  
 Db 679 AGGAGCATACCCAGGCTTTCATTAATCCTTTCTGTTACACCTAATGATATCTTGGGG 620

QY 452 gcgggattcaactgcaagaagttactgagatcgctggtcccaagggttggtaaac 511  
 Db 619 GTGGACTGCCCTTAATGAAACACAGAAATTTGGTGCACACAGTGTGGAAAAACAC 560

QY 512 aactgggattcaactgcaagaatgatacaatccagtggaatgtggtcctgtgtg 571  
 Db 559 AATTATGATGACAGTTCGACAGTAGATGACACAAAGGTTTGGAGGAGTGGCAG 500

QY 572 ggaagcattatataagttacaggggagttcattcattgttgaagtgctacccagatg 631  
 Db 499 GTGAACAGCTTTTATTGATACAGAGGAAATTTATGTTGATGAGTGTAGACCTTG 440

QY 632 ctgaagggtgtattaggagacatactg 657  
 Db 439 CTACTGCGCTGCATTTCAGACACCTTCAG 414

RESULT 13  
 BE390908 658 bp mRNA EST 21-JUL-2000  
 LOCUS 601283055F1 NIH\_MGC\_44 Homo sapiens cDNA clone IMAGE:3605172 5',  
 DEFINITION mRNA sequence.

ACCESSION BE390908  
 VERSION BE390908.1 GI:9336273  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 658)  
 NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>.  
 Tumor Gene Index  
 Unpublished (1999)  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LDCM252 row: m column: 13  
 High quality sequence stop: 624.  
 Location/Qualifiers  
 1. 658  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="3605172"  
 /clone\_lib="NIH\_MGC\_44"  
 /tissue\_type="endometrium, adenocarcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: uterus; Vector: pOT87; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCAAGG(6). Library constructed by Ling Hong  
 in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 183 a 134 c 163 g 178 t  
 ORIGIN

Query Match 6.6%; Score 98; DB 167; Length 658;  
 Best Local Similarity 60.5%; Pred. No. 3.6e-15;  
 Matches 161; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 392 atgagctgcacagaacacacactcactcgtctcgtcgaactcaatgacatactgtg 451  
 Db 222 AGGAGCATACCCAGGCTTTCATTAATCCTTTCTGTTACACCTAATGATATCTTGGGG 281

QY 452 gcgggattcaactgcaagaagttactgagatcgctggtcccaagggttggtaaac 511  
 Db 282 GTGGAGTGCCTTAAATGAAACACAAATTTGGTGCACACAGGTTGGAAAAACAC 341

QY 512 aactgggattcaactgcaagaatgatacaatccagtggaatgtggtcctgtgtg 571  
 Db 342 AATTATGATGACAGTTCGACAGTAGATGACACAAAGGTTTGGAGGAGTGGCAG 401

QY 572 ggaagcattatataagttacaggggagttcattcattgttgaagtgctacccagatg 631  
 Db 402 GTGAACAGCTTTTATTGATACAGAGGAAATTTATGTTGATGAGTGTAGACCTTG 461

QY 632 ctgaagggtgtattaggagacatactg 657

[illegible]

High quality sequence stop: 630.

FEATURES	Location/Qualifiers
source	1. .683

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3630461"
/clone_lib="NIH_MGC_21"
/clone_type="Choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/node="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies). "

```

[illegible]

RESULT	15
BGI77028	
LOCUS	
DEFINITION	BGI77028 778 bp mRNA EST 06-FEB-2001 60233383cft1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4419512 5' , mRNA sequence.
ACCESSION	BGI77028
VERSION	BGI77028.1 GI:12683731
KEYWORDS	EST.
SOURCE	human. Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 778)
REFERENCE	NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
AUTHORS	Contact: Robert Strausberg, Ph.D.
TITLE	
JOURNAL	
COMMENT	

Plate: LLAM10155 row: h column: 09  
High quality sequence stop: 710.

FEATURES	Location/Qualifiers
source	1. .778

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4419512"
/clone_1id="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph. Vector: PCMV-SpORf6; Site_1: NotI; Site_2: SalI. Cloned unidirectionally; oligo-df primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies Note: this is a NIH_MGC Library."

```

BASE COUNT	216 a	151 c	200 g	211 t
ORIGIN				
Query Match	6.6%	Score 98;	DB 174;	Length 778;
Best Local Similarity	60.5%;	Pred. No. 3,8e-15;		
Matches 161;	Conservative 0;	Mismatches 105;	Indels 0;	Gaps
QY	392 atgagctctcaagaacaacatctactctatcgtgtctgtgtgcaccatgaataacttgctg	451		
DB	299 AGGACCATACCCAGGCGCTTCAATACCTTCTGTGACGACTAGATGATATCTTGGGG	358		
QY	452 ggggagttcactgcgaagaagttactagatcgtgtgcgcccaaggggttggtaaacctc	511		
DB	359 GGGGAGTGCCCTTAATGAAAAACAACAGAAATTTGTGGTCACACGAGTGTGGAAAAACAC	418		
QY	512 aactggagattcaactagcaactcaatgtcaacatccagctggaatgtgtgtgacctgtgtg	571		
DB	419 AATTATGATGACGTTGGCAGTAGTAGTGCGCATACCAAGATGTTTGGAGGAGTGGCAG	478		
QY	572 ggaagaagcttatatagatacagaagaaggcagttcatcgttgtgaacgtgtctaccagattg	631		
DB	479 GTGAGCAGATTTTATTTATGATACAGGAGCAAGTTTATGCTTGATAGAGTGATGGTAGACCTTG	538		
QY	632 ctgaagggtgtattagggacatctg	657		
DB	539 CTACTGCGCTGCATTCAGCACCTTAG	564		

Tue Oct 2 11:16:26 2001

us-09-537-654-1.rst

Page 12

Job time: 4425 sec

---



QY 245 NRLLHNGNERYAHLDKSPSPVAPAYATGKIRDAVSSNHRARY 293  
 DB 315 NRVLTYWNGDERAYIDKSPSPVAPASASYTVTSRGLRNS-SSSSKRVKM 362

RESULT 2  
 09FT75 PRELIMINARY; PRT: 363 AA.

AC 09FT75;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE RAD51C PROTEIN.  
 GN RAD51C.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Sland N.;  
 RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ299424; CAC14091.1; -  
 SQ SEQUENCE 363 AA; 40213 MW; E21D09BC0A79FA7A CRC64;

Query Match 66.0%; Score 1010; DB 10; Length 363;  
 Best Local Similarity 64.4%; Pred. No. 4.9e-76;  
 Matches 186; Conservative 57; Mismatches 40; Indels 6; Gaps 2;

QY 5 SSSRNGPQQKYVAGNANDMFSDLSQKHITTSGLNDIIGGICKEVTEIGVPGV 64  
 DB 80 NSRRS-----LNGAKNADMDLHEESLPRTTSCSDLNILGGISCRDVTIEIGVPEI 134  
 QY 65 GKTLQIGLAINVQIPVECGGLGKAVYIDTEGSEFVERVYIABGCIROLLEHPPHSHE 124  
 DB 135 GKQIGIGLSTVNVQIPRECGGLGKAIYIDTEGSEFVRKALQIAECVDEMEETGYMKR 194  
 QY 125 KSSSVQKQLOPEFLADYIFRCSTEQIATVINYMEKRLREHKVRYIIDSVFHFHQ 184  
 DB 195 HFGANQVQKPELIDENIEFYRCSTEQIALVNHLEKTSISKDKVYIVDSITFHFHQ 254  
 QY 185 DFEIDLARTFVLSGLSKLTKIKATNYLAVLLNQVTFTEGSPQLTALDSDMSHSC 244  
 DB 255 DYDDLAGQRTFVLEMLAKFKMLAKRSLAVLLNQVTFTEGSPQLTALDSDMSHSC 314  
 QY 245 NRLLHNGNERYAHLDKSPSPVAPAYATGKIRDAVSSNHRARY 293  
 DB 315 NRVLTYWNGDERAYIDKSPSPVAPASASYTVTSRGLRNS-SSSSKRVKM 362

RESULT 3  
 02144 PRELIMINARY; PRT: 332 AA.

AC 02144;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE PUTATIVE DNA REPAIR PROTEIN (RAD57).  
 GN FAL23.21.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Rounsley S.D.; Lin X.; Ketchum K.A.; Crosby M.L.; Brandon R.C.;  
 RA Sykes S.M.; Mason T.M.; Kerlavage A.R.; Adams M.D.; Somerville C.R.;

RA Venter J.C.;  
 RL Submitted (Nov-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC002387; AAB82635.1; -  
 DR InterPro: IPR001553; -  
 SQ SEQUENCE 332 AA; 36864 MW; A7F058BFD54B8A66 CRC64;

Query Match 64.6%; Score 988; DB 10; Length 332;  
 Best Local Similarity 66.5%; Pred. No. 2.9e-74;  
 Matches 183; Conservative 51; Mismatches 39; Indels 2; Gaps 2;

QY 19 AQNAMDPSDELQKHITTSGLNDIIGGICKEVTEIGVPGVQKQIGLQIALINQ 78  
 DB 59 AKNAMDMLHEESLPRTTSCSDLNILGGISCRDVTIEIGVPGVIGKQIGLQIALINQ 118  
 QY 79 IPVECGGLGKAVYIDTEGSEFVERVYIABGCIROLLEHPPHSHESSSVQKQLOPER 138  
 DB 119 IPRECGGLGKAIYIDTEGSEFVRKALQIAECVDEMEETGYMKRFOANQVQKPEDI 178  
 QY 139 LADYIFRICSTEQIATVINYMEKRLREHKDVRIIVDSYTFHFRODFEDLARTVLSG 198  
 DB 179 LENIFEFRCVCTEQIALVNHLEKTSISKDV-VIVDSITFHFRODYDDLAGRTVLS 237  
 QY 199 LSLKMLKIAKTYNLAVLLNQVTFTEGSPQLTALDSDMSHSCNRLILHNGNERYA 258  
 DB 238 MALKFMKIAKRFSLAVLLNQVTFTEGSPQLTALDSDMSHSCNRLIYWNQDERYA 297  
 QY 259 HLDKSPSPVAPAYATGKIRDAVSSNHRARY 293  
 DB 298 YIDKSPSPVAPASASYTVTSRGLRNS-SSSSKRVKM 331

RESULT 4  
 09FEY1 PRELIMINARY; PRT: 319 AA.

AC 09FEY1;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE 319AA LONG HYPOTHETICAL RADA PROTEIN.  
 GN APE0119.  
 OS Aeropyrum pernix.  
 OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;  
 OC Aeropyrum.  
 OC NCBI\_TaxID=56636;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K1;  
 RX MEDLINE=99310339; PubMed=10382966;  
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,  
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankel A., Kosugi H.,  
 RA Hoshoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
 RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,  
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic hyper-thermophilic  
 RT crenarchaeon, Aeropyrum pernix K1."  
 RT DNA Res. 6:83-101(1999).  
 RL EMBL: AP000058; BAA79030.1; -  
 DR InterPro: IPR001553; -  
 DR InterPro: IPR003593; -  
 DR PRINTS: PR00142; RECA.  
 DR SMART: SM00382; AAA.1.  
 SQ SEQUENCE 319 AA; 35320 MW; 714F88859D619E25 CRC64;

Query Match 27.3%; Score 417; DB 1; Length 319;  
 Best Local Similarity 36.0%; Pred. No. 7.8e-27;  
 Matches 102; Conservative 61; Mismatches 86; Indels 34; Gaps 6;

QY 13 QKYVSGA-----QANAMDPSDELQKHITTSGLNDIIGGICKEVTEIGVPGV 64  
 DB 56 QKIVDAAREALINDFKFAYIDKTESMNIKITTSRNLDELGGIGTETLFGFGS 115



DR InterPro: IPR001553; -  
 DR InterPro: IPR003593; -  
 DR SMART: SM00382; AAA; 1.  
 SO SEQUENCE 349 AA; 38399 MW; 7269DE1D67AB039 CRC64;

Query Match 22.6%; Score 346; DB 1; Length 349;  
 Best Local Similarity 32.1%; Pred. No. 7e-21;  
 Matches 96; Conservative 54; Mismatches 105; Indels 44; Gaps 8;

QY 5 SGRNPGQKYYVSGAONAMD---MFSDELSOK-----HITSGDLNDILGGIHCKEY 55  
 DB 74 AGISEGTALKIIOAKRANKLTGEMRADYELKRRATIGRISTGSLDLGGIETQAI 133  
 QY 56 TEIGVGVGVKQTOLGIOLAINVOIPVECGGLGKAVYIDTESFWERYQIAEGCIRDI 115  
 DB 134 TEYVGFEGSGKTOLAHITLAVMQLPPEEGGLNSVIMIDTENTFRERIREIAQ----- 187  
 QY 116 LEHFPHSHEKSSSVOKOLOPERFLADYYFRICSTEQIAYINMEKFLRE---HKDVR 171  
 DB 188 -----NRGLDPEVLKHIYVARAFNSNHOMLVOQAEDKIKELINTEDRPVK 233  
 QY 172 IVIIDSVTFFHRODF---EDLALRTVLSGLSLKIMKIATYTNLAVLLNOVTTK----F 224  
 DB 234 LLIIVDSLHSFRSEYIGRGLAERQOKLAKHLADLRLNLDIAVFNVOVARDAF 293  
 QY 225 TEGSFQTLALGDSWSHSCNRLILH-WGNERYAHLDKSPSPVAPYATVGKIRD 282  
 DB 294 GDTFRPIG---GHILHSAITLRYLKRKGKRIALIDAPHLPEGEAVFITEKGIED 349

RESULT 8  
 ID Q9V233 PRELIMINARY; PRT; 356 AA.  
 AC Q9V233;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE DNA REPAIR PROTEIN RAD51 (RADA).  
 GN RADA OR PAB0164.  
 OS Pyrococcus abyssi.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 OX NCBI\_TaxID=29282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ORSAY;  
 RA Hellig R.;  
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."  
 RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AJ248283; CAB49165.1; -  
 DR InterPro: IPR001553; -  
 DR InterPro: IPR003593; -  
 DR PRINTS: PR00142; RECA.  
 DR SMART: SM00382; AAA; 1.  
 SO SEQUENCE 356 AA; 39136 MW; 96270CE1D64F6CEE CRC64;

Query Match 22.5%; Score 345; DB 1; Length 356;  
 Best Local Similarity 32.4%; Pred. No. 8.7e-21;  
 Matches 97; Conservative 51; Mismatches 107; Indels 44; Gaps 8;

QY 5 SGRNPGQKYYVSGAONAMD---MFSDELSOK-----HITSGDLNDILGGIHCKEY 55  
 DB 81 AGISEGALKIIOAKRANKLTGEMRADYELKRRATIGRISTGSLDLGGIETQAI 140  
 QY 56 TEIGVGVGVKQTOLGIOLAINVOIPVECGGLGKAVYIDTESFWERYQIAEGCIRDI 115  
 DB 141 TEYVGFEGSGKTOLAHITLAVMQLPPEEGGLNSVIMIDTENTFRERIREIAK----- 194  
 QY 116 LEHFPHSHEKSSSVOKOLOPERFLADYYFRICSTEQIAYINMEKFLRE---HKDVR 171  
 DB 195 -----NRGLDPEVLKHIYVARAFNSNHOMLVOQAEDKIKELINTDQPKV 240

QY 172 IVIIDSVTFFHRODF---EDLALRTVLSGLSLKIMKIATYTNLAVLLNOVTTK----F 224  
 DB 241 LLIIVDSLHSFRSEYIGRGLAERQOKLAKHLADLRLNLDIAVFNVOVARDAF 300  
 QY 225 TEGSFQTLALGDSWSHSCNRLILH-WGNERYAHLDKSPSPVAPYATVGKIRD 282  
 DB 301 GDTFRPIG---GHILHSAITLRYLKRKGKRIALIDAPHLPEGEAVFITEKGIED 356

RESULT 9  
 ID Q9NCP0 PRELIMINARY; PRT; 236 AA.  
 AC Q9NCP0;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE DNA REPAIR PROTEIN RAD51 (FRAGMENT).  
 GN RAD51.  
 OS Giardia lamblia (Giardia intestinalis).  
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardiae; Giardia.  
 OX NCBI\_TaxID=5741;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20293306; PubMed=10833194;  
 RA Inagaki Y., Doolittle W.F.;  
 RT "Evolution of the eukaryotic translation termination system: origins of release factors."  
 RL Mol. Biol. Evol. 17:882-889(2000).  
 DR EMBL; AF198107; AAF74403.1; -  
 DR InterPro: IPR001553; -  
 DR InterPro: IPR003593; -  
 DR SMART: SM00382; AAA; 1.  
 FT NON\_TER 1  
 SO SEQUENCE 236 AA; 25261 MW; CEFA7412568F6D08 CRC64;

Query Match 21.6%; Score 330.5; DB 5; Length 236;  
 Best Local Similarity 35.9%; Pred. No. 8e-20;  
 Matches 92; Conservative 39; Mismatches 94; Indels 31; Gaps 6;

QY 37 TGSGLNDILGGIHCKEYTEIGVGVKQTOLGIOLAINVOIPVECGGLGKAVYIDTE 96  
 DB 1 SCADFNALGSGVETMSITVEFGEFRGKTQLCHITLAVNQLVSGGGGKTYIDTE 60  
 QY 97 GSFWEERYQIAEGCIRDIIEHFPHSHEKSSSVOKOLOPERFLADYYFRICSTEQIAV 156  
 DB 61 GTFREKXVPIAE-----RFGINPKKALDINIVARYTTHQOIEC 100  
 QY 157 INYMEKFLREHKDVRIYIIDSVTFFHRODF---DLALRTVLSGLSLKIMKIATYTNLA 213  
 DB 101 ITALPKLWENQ-FSLYIVDSLTLALFRVDFGRGELADROOKLGQHLGLAKLADFNLA 159  
 QY 214 VLLNOVTTKFTESFQTL-----ALGDSWSHSCNRLILH-WGNERYAHLDKSPSLP 267  
 DB 160 VFTVQVNAQ-VDGAAMETADPKKPIGSHILAHASTIRLYLKRKGDTFRAKIDSPSLA 218  
 QY 268 VASAPYATVGKIRDA 283  
 DB 219 EGEASTIAEGTIDA 234

RESULT 10  
 ID Q9DGC3 PRELIMINARY; PRT; 342 AA.  
 AC Q9DGC3;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE DMCL.  
 GN DMCL.  
 OS Cynops pyrrhogaster (Japanese common newt).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Caudata; Salamandroidae; Salamandridae; Cynops.  
 OX NCBI\_TaxID=8330;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-TESTIS;  
 RA Yezawa T., Yamamoto T., Nakayama Y., Hamada S., Abe S.;  
 RT "Conversion from mitosis to meiosis: morphology and expression of PCNA  
 and DMC1 during newt spermatogenesis."  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB041944; BAB16892.1;  
 SQ SEQUENCE 342 AA; 37981 MW; 868E002AEAF0BEEB CRC64;

Query Match 20.8%; Score 318; DB 13; Length 342;  
 Best Local Similarity 33.5%; Pred. No. 1,4e-18;  
 Matches 90; Conservative 44; Mismatches 101; Indels 34; Gaps 8;

OY 29 ELISK-----HTTGGDNDLILGGIHCKEVTEIGVGKTOGLQIALINVOIPVEEC 83  
 DB 92 EYSEKRMVHTTGGSOEFDKLGGLGIESMAITTFEFPRTGKQLSHLTCTVRLPFTD 151  
 OY 84 GGLGKAVYIDTEGSMFVERVYQIAEGCIRDILEHFPHSHEKSSSVOKOLOPERFLADY 143  
 DB 152 GYTGKVIPTDENTFRDR-----LRDIADRFVSVDH-----DAVDNL 191  
 OY 144 YFRICSTEQIAVINIM-EKFLREKDVRIYIDSVFHRDPE---DLALTRVLSGL 199  
 DB 192 YARAVTSEHOMELDYAAKHEBAGIFKLLIDISIMALFRVDSGKGLAEKQKLAQM 251  
 OY 200 SLKLMKIAKTYNLAIVLLNOVTT-KTEGSPQLTAL---GDSWSHSCNRLILH-WNGN 254  
 DB 222 LARLOKISEEVNAVAVYNTQAPDGAAMSQADPKPKIGIHILAHSTRISLRKRG 311  
 OY 225 ERYAHLKSPSLPVASAPYAVTGKIRDA 283  
 DB 312 LRIAKIYDSPMPENEAFTAGIGIDA 340

RESULT 11

O61128 PRELIMINARY; PRT; 364 AA.  
 ID O61128  
 AC O61128  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE DMC1 HOMOLOG.  
 GN DMC1.  
 OS Leishmania major.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FRIEDLIN;  
 RA McKean P.G., Benson F.E.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF062380; AAC16335.1;  
 DR HSSP: P03017; 2REB.  
 DR InterPro: IPR001553;  
 DR InterPro: IPR003593;  
 DR SMART: SM00382; AAA; 1.  
 SQ SEQUENCE 364 AA; 39527 MW; A5475644CE8AFBE5 CRC64;

Query Match 20.7%; Score 316; DB 5; Length 364;  
 Best Local Similarity 34.5%; Pred. No. 2.3e-18;  
 Matches 90; Conservative 49; Mismatches 88; Indels 34; Gaps 8;

OY 35 ITTGGDNDLIL-GGIIHCKEVTEIGVGKTOGLQIALINVOIPVECGLGKAVYI 93  
 DB 126 ISTGSAIDQLLGGGIERSTTEAFGERRTKTOIGHTLCTQCLPLEMGKNAKAVY 185  
 OY 94 DTGGSFVERVYQIAEGCIRDILEHFPHSHEKSSSVOKOLOPERFLADYIFRICSYEQ 153

DB 186 DTGTFPRPRIRPIAEFCMD-----SNSV-----LDNIIYARAYTHEQ 225  
 OY 154 IAVINMEFLAEKDVRIYIDSVFHRDPE---DLALTRVLSGLSKIMKIATY 210  
 DB 226 AHLISVAKMAEDQ-FSLVVDSTALFRVDSGKGLAEKQKLAQLSOLITAEF 284  
 OY 211 NLAIVLLNOVTTKFTGSPQLTIA-----LGDMSHSCNRLILH-WNGENRYAHLDS 263  
 DB 285 NIAVITTYNOVS--DTGSMVADPKPKVGGHILAHSTRISLRKGGDQRCVKIDS 342  
 OY 264 PSLPVASAPYAVTGKIRDAV 284  
 DB 343 PSLPELECYTISEGIIIDAV 363

RESULT 12  
 O90H11 PRELIMINARY; PRT; 340 AA.

ID O90H11  
 AC O90H11  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE DJ199H16.1 (DMC1 (DOSAGE SUPPRESSOR OF MCK1, YEAST HOMOLOG) MEIOSIS-SPECIFIC HOMOLOGOUS RECOMBINATION PROTEIN (HIM15))).  
 GN DMC1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Blakey S.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF022320; CAB45656.1;  
 DR HSSP: P03017; 2REB.  
 DR InterPro: IPR001553;  
 DR InterPro: IPR003593;  
 DR SMART: SM00382; AAA; 1.  
 SQ SEQUENCE 340 AA; 37681 MW; 040A6E4CF1EBFA2 CRC64;

Query Match 20.5%; Score 313; DB 4; Length 340;  
 Best Local Similarity 32.7%; Pred. No. 3.7e-18;  
 Matches 89; Conservative 45; Mismatches 98; Indels 40; Gaps 8;

OY 29 ELISK-----HTTGGDNDLILGGIHCKEVTEIGVGKTOGLQIALINVOIPVEEC 83  
 DB 90 EYSEKRMVHTTGGSOEFDKLGGLGIESMAITTEAFGERRTGKQLSHLTCTVRLPFG 149  
 OY 84 GGLGKAVYIDTEGSMFVERVYQIAEGCIRDILEHFPHSHEKSSSVOKOLOPERFLADY 143  
 DB 150 GYTGKVIPTDENTFRDR-----LRDIADRFVSVDH-----DAVDNL 189  
 OY 144 YFRICSTEQIAVINIM-EKFLREKDVRIYIDSVFHRDPE---DLALTRVLSGL 199  
 DB 190 YARAVTSEHOMELDYAAKHEBAGIFKLLIDISIMALFRVDSGKGLAEKQKLAQM 249  
 OY 200 SLKLMKIAKTYNLAIVLLNOVTTKFTGSPQLTIA-----GDSWSHSCNRLILH-W 251  
 DB 250 LSRLOKISEEVNAVAVYNTQAPDGAAMSQADPKPKIGIHILAHSTRISLRKRG 306  
 OY 252 NGENRYAHLDSPSLPVASAPYAVTGKIRDA 283  
 DB 307 RGLRIAKIYDSPMPENEAFTAGIGIDA 338

RESULT 13  
 O44246 PRELIMINARY; PRT; 357 AA.

ID O44246  
 AC O44246  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)





